

Short-circuit to be avoided by bioethics committees

A decision by the French national bioethics committee to recommend a ban on human cloning for reproductive purposes sets a dangerous precedent, pre-empting a much-needed public debate.

The human cloning debate has revealed the growing political importance of bioethics committees in the regulation of new technologies. Confronted with public concern about human cloning, the instant reflex of political leaders worldwide has been to call on the nearest bioethics committee for advice.

The political weight now being given to such committees (see pages 323, 324) requires a re-examination of both their operation and their place in the democratic process of decision-making. Such committees are ultimately a small group of unelected individuals. The danger to democracy is that although such committees have no legislative powers, they often exert considerable influence on legislators unfamiliar with complex scientific and ethical issues.

The French committee's ruling should serve as a reminder of this danger. To its credit, the committee has produced an excellent 49-page overall review of cloning, with the scientific issues set out in a form which even the least scientifically literate members of the French National Assembly could grasp. Where the committee errs is in the nine pages devoted to ethical aspects of human cloning. Members of the committee concede that all the various ethical arguments it puts forward could be individually contested, but maintain that, taken together, they support a ban on cloning. That conclusion remains contentious. But the essential point is that, in deciding to take a moral position on human cloning, the committee has gone beyond its role, which should simply be to clarify the debate and promote informed public debate.

The big risk is that politicians will use such ready-made solutions as an excuse to abdicate their duty to shape their legislative and constitutional provisions on bioethics to the religious, moral and cultural traditions of society, as well as to the dominant streams of contemporary thought. For example, Jacques Chirac, the French president, wasted no time in using the committee's report to support the popu-

lar political stance of calling for an international ban on the use of human cloning for reproduction.

Questions about the democratic legitimacy of bioethics committees also seem to have escaped Chirac, who proudly declared that the ethics committee had admirably fulfilled its role of telling the government "what should be done" in the face of a bioethical problem. But giving ethics committees this role short-circuits the democratic process by pre-empting wide public debate.

On issues such as human cloning, where emotions run high, ethical committees have an even greater responsibility to provide a measured response. Indeed, in cloning and other such situations, ethics committees are often the last rampart against a stampede of irrational and opportunistic political judgement. A report by the World Health Organization's working group on cloning argues that opposition to human cloning "has prompted legislators and other policy-makers to act out of 'moral panic' rather than from considered deliberation".

The International Bioethics Committee of Unesco is scheduled this week to release an opinion on human cloning. Its job has been made awkward by the premature declaration of Federico Mayor, Unesco's director general, that "human beings should not be cloned under any circumstances", and his limited remit that the committee should verify whether the draft Unesco "universal declaration on the human genome and human rights" outlaws cloning.

In fact, the committee is deeply divided. It should also ask Mayor for more time and for an enlarged remit that would allow it to produce a more comprehensive account of the complex issues involved. It would be doing a service to public debate if it were then to explain the basis of the divergent opinions. Meanwhile, bioethics committees elsewhere should ensure that they avoid becoming undemocratic bastions of premature and forced consensus. □

A need for endorsement?

Science studiers and their scientist critics should concentrate on constructive engagement.

A dispute at the Institute of Advanced Study at Princeton about a decision not to appoint a historian of science to a position in social sciences (see page 325) may reflect the grey area that can exist between objective assessment of research quality and vehement opposition to ideas. A central issue is said to have been the way in which sociologists and others analyse the social dimensions of science through the discipline known as 'science studies'. More specifically, hostility that has developed among some scientists against the part of these studies known as the 'sociology of scientific knowledge' is being blamed for the failure of the proposed appointment of an individual sympathetic to this approach to generate the required support.

That hostility is understandable. The status of scientific knowledge is approached by the two sides very differently. Scientists tend to see this knowledge as a form of truth embedded in nature, waiting to

be revealed by their experimental and theoretical skills. Some sociologists of science, in contrast, prefer to present this 'truth' as the product of negotiations between scientists — hence the criticism from natural scientists that they present a 'relativist' or 'idealist' view of scientific knowledge.

The Princeton decision, and debates elsewhere in the US academic community, reflects the fact that careers are now at stake amid heated — sometimes vitriolic — exchanges generated in what has come to be known as the Science Wars (see Briefing, pages 331–335). Scientists, rightly suspicious of relativism, should nevertheless be circumspect in pursuing this debate. If, as a result of forthcoming meetings between scientists and sociologists of science, the former come to appreciate the virtues of many aspects of science studies, they should say so, loudly and clearly. Good scholarship is at risk in the current climate. □



Clinton sketches out his 'ethical guideposts' for modern biology

[WASHINGTON] In the most prominent speech of his presidency on research-related issues, President Bill Clinton called last Sunday (18 May) for Americans to remember that "science is not God", adding: "Our deepest truths remain outside the realm of science."

In a speech that has been noticed mainly for its call for development of an AIDS vaccine, Clinton also spoke widely on scientific ethics. He urged Congress to pass legislation banning genetic discrimination by health insurers. Referring to the recent cloning of a sheep, he said that "we should resist the temptation to replicate ourselves".

"Science has no soul of its own," said Clinton. "It is up to us to determine whether it will be used as a force for good or evil."

Clinton's call for an AIDS vaccine was accompanied by an announcement that a new centre devoted to AIDS vaccine research will be established at the National Institutes of Health (NIH). But as he did not increase the \$148 million AIDS vaccine budget at NIH, the announcement has been met with some scepticism.

One highly placed AIDS scientist dismissed it as "pure politics". But others, such as Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases (NIAID), welcomed the centre and called the ten-year goal an appropriate symbolic pledge. Harold Varmus, director of NIH, described the goal as "realistic".

In his speech, delivered as a commencement address at Morgan State University in Baltimore, Maryland, Clinton outlined four ethical "guideposts" for the coming "century of biology". These were: that science should be used to produce a better life for all, and not the few; that advances should not be used to discriminate and, to this end, that Congress should ban genetic discrimination by health insurers; that new technologies should not be allowed to compromise individual privacy, including medical records privacy; and that American science should "never [confuse] faith in technology with faith in God".

Referring to the government's Tuskegee Syphilis Study, in which 399 black men were left untreated, and to Cold War radiation experiments on US citizens, Clinton vowed that America would "not go back to those awful days in modern disguise".

Clinton dismissed some scientists' questions about whether the formidable obstacles can ever be overcome to produce an AIDS vaccine. "It is no longer a question of whether we can develop an AIDS vaccine, it is simply a question of when." He added that he would appeal to other countries to join



Clinton: getting to grips with AIDS research?

the US effort next month in Denver, at a summit of leading industrialized nations.

His pledge has been warmly welcomed by organizations addressing the global AIDS epidemic, which fear that the availability of protease inhibitors to treat AIDS patients in wealthy countries may undermine support for vaccine research (see *Nature* 382, 101; 1996). "This is the best news that the developing countries could have," says Jose Esparza, a vaccine adviser at the United Nations Programme on AIDS in Geneva, Switzerland.

Clinton also vowed to "challenge" the pharmaceutical industry to invest more money in AIDS vaccine development. This provoked a response from the president of the Pharmaceutical Research and Manufacturers of America that "HIV vaccine development depends more on scientists' tenacity and brilliance in the laboratories, not on politician's words".

But William Paul, director of the NIH's Office of AIDS Research, says that drug com-

pany executives have "assured" NIH that "if we can show the way [with the science] they are prepared to follow through" with vaccine development.

Several AIDS groups were supportive of Clinton's call. Gary Rose of the Washington-based AIDS Action Council calls it a "hopeful" step. "We agree with the president that [vaccine development] should be one of the major targets now," says Arthur Ammann, president of the American Foundation for AIDS Research. "We've got to look at where the epidemic is going."

But others argue that Clinton was fed meaningless declarations by his Advisory Council on HIV/AIDS, which proposed the ten-year goal in April. "These are people who don't deal with the scientific issues. They believe that the wish is the father of the deed," says a prominent AIDS researcher, maintaining that, without more money, the declaration is purely cosmetic.

According to Varmus, despite the lack of new money accompanying Clinton's announcement, "there are ways" to make adequate resources available for the centre. Fauci says the president's declaration will give the NIH leverage in coming years to land healthy AIDS vaccine research budgets even as other government programmes shrink.

The NIH vaccine research centre will consolidate AIDS vaccine research now located at the NIAID and the National Cancer Institute. Varmus hopes to house 50 to 100 core scientists and staff in one facility within one to two years, and involve other NIH scientists in a larger, 'virtual' centre without walls.

Meredith Wadman

Bioethics commission gets extended term

[WASHINGTON] Prompted largely by public concern about humans used in experiments, President Bill Clinton last week extended by two years the life of the national bioethics commission, which is to report to him on cloning at the end of this month (see *Nature* 387, 217; 1997).

"The need for this commission is clear," Clinton said at a ceremony at the White House during which he announced that he was using an executive order to extend the life of the National

Bioethics Advisory Commission until October 1999. The commission would otherwise have ceased work in October this year.

The brief mandate and thin resources of the 18-member commission had come under fire from Republicans at a congressional hearing two weeks ago. There, and in news reports since, the US system for the protection of human subjects has been portrayed as inadequate.

"We must be able to call on the thoughtful, collective

wisdom of experts and community representatives to find ways to further strengthen our protection for subjects in human research," Clinton said. Part of the commission's mandate is to focus on the welfare of human research subjects.

The chairman of the commission, Harold Shapiro, who is president of Princeton University, said of Clinton's announcement: "The president has assured me all along that this was going to happen. It's nice to get it done."

M. W.

Calls for human cloning ban 'stem from ignorance'

[PARIS] The recent flood of calls for an international ban on the use of human cloning techniques for reproduction appear, at least in some quarters, to be slowly giving way to recognition that such practices may be justified in certain circumstances.

There is also a growing feeling that advocates of a ban have so far failed adequately to substantiate their claims that cloning would be either harmful or unethical.

A report published this month by a working group on cloning set up by the World Health Organization (WHO), for example, argues that much of the opposition to human cloning stems from "science fiction accounts" which have resulted in "fear and ignorance on the part of the public".

The report says that this has prompted legislators and other policy-makers to act from 'moral panic' rather than considered deliberation. It says that the cloning debate involves many issues that still need to be discussed in detail, and concludes that introducing an immediate international ban on cloning would be "unwise and counter-productive".

The report adds that a ban or moratorium may be "most incautious", as a hasty prohibition could result in loss of actual and potential benefits. But its conclusion that WHO should proceed more carefully appears to have gone unheeded by the organization's General Assembly, a political body representing WHO's member states. Last week, the assembly adopted a resolution affirming that "the use of cloning for the replication of human individuals is ethically unacceptable and contrary to human integrity and morality".

The assembly's resolution follows similar calls for an international ban from political leaders such as Jacques Chirac, the French president, and Jürgen Rüttgers, Germany's research minister (see *Nature* 387, 111; 1997). Such opposition appears to be based on a perception that the benefits of human cloning would be few, whereas the risks of abuse could be large. But critics argue that the benefits are being underestimated and the risks overstated.

"A sense of proportion is needed," says David Shapiro, executive secretary of the UK Nuffield Council on Bioethics, who points out that there is plenty of time for debate, given the enormous technical hurdles to be overcome before human cloning could even be authorized on safety grounds.

Any demand for cloning would also be relatively small and could be tightly regulated, predicts Shapiro, arguing that, from an ethical point of view, the technique is not

significantly different from many accepted forms of medically assisted procreation.

John Robertson, professor of law at the University of Texas in Austin, last week told a meeting of the International Bioethics Committee of the United Nations Educational, Scientific, and Cultural Organization that "initial repugnance has given way to the recognition that there may be some benefits to infertile couples and others from human cloning".

He added that there was also a growing awareness that "the harms alleged to flow from cloning are too vague and speculative to justify a ban on all possible uses of cloning or on cloning research".

Indeed, the widespread political opposition to cloning is causing concern among many scientists and bioethicists. Considered discussion is falling victim to emotion and politics, says Giuseppe Benagiano, director of WHO's programme on research on human reproduction.

"The president or prime minister stands up and says you have to ban human cloning; everybody applauds, he gets more votes, and the arguments play no role," says Benagiano. The fact that the assembly also requested WHO to assess fully the ethical, scientific and social implications of human cloning is "the only thing that is keeping the door open", he adds.

David Griffin, the secretary of WHO's

working group on cloning, says the arguments are highly complex, while attitudes to applications of human cloning differ widely between cultures. He argues that a "lack of information" has obscured the debate.

The WHO report points out, for example, that while producing a cloned twin child for spare organs and tissues would clearly be ethically unacceptable, individuals in some cultures might not object to producing clones of early human embryos as a source of spare parts — via the production of embryonic stem cells, perhaps (see *Nature* 387, 218).

Infertile couples are also likely to be a source of demand for cloning. In couples where both partners lack gametes, cloning could provide an alternative to the current practice of embryo donation. It could also be used by couples where the male partner lacks gametes, as it might be considered preferable to using sperm from donors.

Couples undergoing *in vitro* fertilization may also wish to use cloning to generate extra embryos and increase the chances of fertilization in cases where the female partner has few oocytes. Robertson recently told a panel of the US National Bioethics Advisory Commission that if this was essential to reproducing, the couple might have a legal right to the technique under US law, as it would fall under the fundamental freedom to reproduce.

Robertson argues that for most realistic applications of cloning, it is "difficult to see harm for either children, families or society". A ban on cloning for reproduction or on research that might lead to it, "is imprudent and unjustified," he asserts. "Science fiction should not guide science policy." **Declan Butler**

Unesco bioethics experts query their remit

[PARIS] Members of the International Bioethics Committee of the United Nations Educational, Scientific and Cultural Organization (Unesco) appear to be divided on the issue of human cloning, on which they are due to produce an interim statement this week.

In particular, many are unhappy about the remit given to the committee by Federico Mayor, director general of the agency, to check whether Unesco's draft "universal declaration on the human genome and human rights" outlaws cloning. Some are also dismayed that Mayor has publicly declared that humans "should not be

cloned" in any circumstances. "It would have been better if Unesco and the director general had waited," says one committee member.

In addition, many members feel that the remit and Mayor's declaration have already tied its hands unnecessarily, arguing that they would have preferred to have carried out a broad assessment of the issues raised by cloning. And several question the wisdom of a ban on human cloning (see above). Some committee members feel it is unlikely that the declaration explicitly bans human cloning as it stands, while several also argue that in any case the proposed declaration is about human

rights and not about specific applications.

Noëlle Lenoir, the committee's chairwoman, has previously stated that the declaration "stops short of including detailed provisions governing medical and research practices [such as genetic tests or gene therapy] on the grounds that doing so would be inappropriate and unworkable within an international context".

Many committee members are also opposed to amending the declaration to ban human cloning explicitly. They fear that tinkering with the declaration risks unravelling the four years of international negotiations it involved. **D.B.**

'Science Wars' blamed for loss of post

[WASHINGTON] Social scientists at the Institute for Advanced Study at Princeton, New Jersey, have decided to stop trying to build up a research capability in 'science studies' after failing to secure the appointment of a historian of science from Princeton University to a faculty position.

The decision has been taken against a background of intense dispute with natural scientists at the institute, who had opposed the appointment of the historian — Norton Wise — because, the social scientists allege, of their hostility to science studies as an academic discipline.

The Institute for Advanced Study is a small but distinguished research school, unconnected to the adjacent Princeton University. Wise's appointment was halted by the institute director, Phillip Griffiths, after a specially appointed panel had split 4:2 in Wise's favour.

The panel consisted of three outside experts in science studies and representatives of each of the institute's other three schools — history, mathematics and natural sciences. It was set up after a similar row six years ago, when the French sociologist Bruno Latour was rejected for the same slot.

In the interim, the institute had agreed to appoint Peter Galison, a historian of science at Harvard University. But Galison eventually turned down the offer because his wife was unable to find a suitable job nearby. The school of social science will now return a \$500,000 grant, intended to fund the position, to the Henry Luce Foundation.

After the panel vote, Griffiths asked for outside advice from, among others, the physicists Steven Weinberg and Gerald

Holton. Weinberg had publicly clashed with Wise in the increasingly acrimonious series of arguments about the study of science, sometimes known as the 'Science Wars' (see Briefing, page 331).

Social scientists at the institute describe this exercise as 'a fishing expedition' deliberately mounted to gather criticism of Wise.

Officials at the institute say that the final decision, first reported in the *Chronicle of Higher Education*, was based purely on Wise's merits as a scholar. Griffiths says: "This was simply an issue of the appointment of an individual — it was not a 'Science Wars' issue at all."

Griffiths declines to discuss these merits specifically, and Ed Witten, the string theorist and one of the panel members who voted against Wise, refuses to comment on the decision. Glen Bowersock, a classical historian and the second 'no', says that it was "a judgement call based on quality", not on Wise's field of study.

According to one outside academic, Wise's publishing record was felt to be weak, both qualitatively and quantitatively. "The scientists are not stupid and they are not ill-willed," the academic says. "What they'd like to have is quality."

But Joan Scott, one of the three current faculty members in the social science school, accuses Wise's detractors of "taking a McCarthyist approach" in the run-up to the decision, and calls Griffiths' denial "absurd".

She argues that Wise was excluded because of his position in the Science Wars debate. "We are now abandoning our attempt to represent science studies at the institute," she says. "We know who the top

Wise: supporters say failed appointment was due to hostility to science studies as a discipline.

people are in the field, and we don't see how we can come up with somebody that will be acceptable on all of the grounds that the institute requires. We've exhausted all the possibilities."

Clifford Geertz, an anthropologist and member of the social science faculty, also refuses to believe that the decision about Wise was based solely on merit. "Wise is a figure of great standing in history of science, as letters from his peers demonstrated," he says, adding that a group of "very determined people" on the natural science faculty had worked to block the appointment.

Wise admits that he has not published much in recent years, saying he has been preoccupied with improving the history of science programme at Princeton University. He says his main concern is that his non-appointment will discourage young historians of science who are already alarmed by what he terms "direct threats" that have been made by some scientists to regain control over science studies. **Colin Macilwain**

Row over subsidies could delay EU Framework programme

[MUNICH] Political disagreement between member states of the European Union threatens to delay the start of the EU's fifth Framework research programme (FP5), due to begin in 1999, by disrupting the programme's approval procedure.

At a meeting of the EU council of research ministers last week, Spain refused to allow a formal decision to be reached on draft proposals for FP5 that have been drawn up by the European Commission until the commission's overall budget beyond 1999 has been agreed.

Spain is concerned that after 1999 the EU's subsidies to its poorer regions, known as 'cohesion funds' and from which Spain benefits significantly, may be severely cut. Its concern is said to be shared by Ireland, Portugal and Greece.

These countries could block progress on FP5 as a way of putting pressure on other EU

members to reverse the cuts. As a result, many national delegations to the council fear that the current, fourth Framework programme may have to be extended by a year to fill the funding gap. The proposals for FP5 require the unanimous approval of the council.

Although Spain's reservations meant that no formal decisions could be taken, last week's council meeting agreed that the scope of the commission's FP5 proposal should be slightly expanded. Officials say they had kept the scope deliberately narrow to ensure a more targeted use of funds.

The commission had proposed three broad thematic programmes, including 16 so-called 'key actions' (see *Nature* 386, 527; 1997). A broad consensus was reached last week that two of the thematic programmes should be split into more manageable components. That would mean separating

life sciences from environmental issues, and separating energy from a general 'industrial technologies' package, creating two new thematic programmes.

Several countries, including Germany, France and the United Kingdom, also pressed for a further thematic programme on transport, but this was rejected by other member states.

The council of research ministers also called on the commission to provide a clearer definition of its proposed role for basic research in FP5. This is not explicitly mentioned in the proposal. Most delegations want assurance that basic research would be funded only in limited and clearly defined areas such as neuroscience or nanotechnology.

In a statement, the council declared its resolve to "make every effort" to enable FP5 to start on time. **Alison Abbott**

HUGO warning over broad patents on gene sequences

[LONDON] The Human Genome Organization (HUGO), the international body that coordinates research into the sequences of the human genome, has confirmed its opposition to the patenting of "short sequences from randomly isolated portions of genes encoding proteins of uncertain functions".

Its statement, echoing concern expressed previously, coincides with news that the US Patent and Trademark Office had provisionally agreed to issue patents on expressed sequence tags (ESTs). The patent office had accepted the argument that the use of such tags as probes to identify specific DNA sequences demonstrates their clear 'utility' (see *Nature* 385, 670 & 386, 747; 1997).

HUGO's Intellectual Property Rights Committee urges the US patent office and others that have adopted a similar position to rescind their decision and, in the meantime, strictly to limit the claims of patents to specified uses. "It would be untenable to make all subsequent innovation in which EST sequences would be involved in one way or other dependent on such patents," the statement says.

At the same time, the committee expresses caution about excessive commitment to the free use of sequence data, implicitly warning against any attempt to impose a blanket ban on the patenting of complete human genes. While reaffirming its view that large-scale sequencing laboratories should make raw sequence data freely available, the committee expresses the hope that this "will not unduly prevent the protection of genes as new drug targets" — arguing that such protection "is essential for securing adequate high-risk investments".

The HUGO statement argues that legislators should seek an international agreement on the introduction of a 'grace period', similar to that which exists in the United States, allowing researchers a period after the publication of their results in which they can still file for patent protection.

The panel that drafted the HUGO statement was chaired by Joseph Strauss, of the Max Planck Institute for Foreign and International Patent, Copyright and Competition Law in Munich. Its members included both industrial scientists, such as Peter Goodfellow of SmithKline Beecham, and academic researchers, such as John Sulston, director of the Sanger Centre in Cambridge, England.

It stresses that "only the policy of rapid publication and free availability of human genomic sequence information will secure further international co-operation of large-scale sequencing centres". □

Africa spearheads bid for strict rules on biosafety

[LONDON] Fifteen African countries unexpectedly seized the initiative last week at a meeting of the United Nations Biodiversity Convention by tabling detailed proposals calling on all signatories to the convention to accept tough regulations governing the use of genetically modified organisms.

Led by Ethiopia, the African proposals call for an all-embracing international biosafety protocol to govern the handling, transfer, use and release of all 'living modified organisms' (LMOs).

The proposals are being opposed by the European Union (EU) and the United States, as they appear to contravene an agreement made last year at the annual conference of the biodiversity convention. At that meeting, participating countries agreed that a biosafety protocol would cover only crossborder transport of LMOs (see *Nature* 382, 384; 1996).

Sateev Seebaluck, principal assistant secretary at the Ministry of Environment and Quality of Life in Mauritius, says Africa did not so much change its mind as exhibit a "delayed reaction" to the original consensus. "African countries were silent, but that did not mean we agreed with the EU position."

The African document was thrashed out last year in the Ethiopian capital Addis Ababa at a meeting of eight countries, including Zambia, Mali, Cameroon and Mauritius, as well as Ethiopia. A further nine countries signed up to the document last week at a meeting of the biosafety protocol working group in Montreal.

The document insists that the protocol should include detailed provisions for liability and compensation payments in the event of harm being caused by the use of LMOs.

Some claim that the African countries have been assisted in drawing up their proposals by at least one Western non-governmental organization; one name being mentioned is that of the Community Nutrition Institute in Washington DC.

African countries are calling for the protocol to take account of the social and economic consequences of biotechnology, such as the loss of revenue for countries whose exports are affected when countries stop importing traditionally grown crops in favour of growing or importing genetically modified ones.

European countries with links to Africa are understood to be trying to persuade African signatories not to adopt tough measures, which they believe could inhibit trade and agriculture. Most EU countries would prefer to leave detailed regulation of genetically modified organisms to national authorities.

Just before last week's meeting in Montreal, officials from French-speaking developed



countries, including France, Canada and Switzerland, organized a two-day conference in the city with counterparts from Francophone Africa. The meeting was designed to help French-speaking African countries keep abreast of developments within the convention, as official UN biodiversity documents are not yet available in French.

But Yoke-Ling Chee, a representative of the environmentalist group Third World Network based in Penang, Malaysia, echoes the views of several observers by saying "there were rumours that the meeting may have been an attempt to lobby African countries to change their positions".

Seebaluck acknowledges that pressure will be put on African countries to modify their positions in the run-up to the next biosafety meeting in October, which is the deadline for submissions of proposed text for the protocol, and when hard negotiations will begin. He points out that South Africa is already said to be uneasy, particularly on the questions of liability and the economic implications of biotechnology.

Significantly, however, developing countries remain divided on the content and scope of the proposed biosafety protocol. Countries in the Far East and Latin America that are keen to grasp biotechnology to enhance trade and agriculture are broadly opposed to tough regulations. Less developed countries are in favour of such measures.

India, which belongs to the former group, has not yet formally shown its hand. But sources close to the Indian government say that it would not want to sign up to a protocol that could impede possible commercial advantages from biotechnology.

Many developing countries that were in favour of strong regulations two or three years ago are said to have privately changed their positions, but to be unprepared as yet to say so in public.

Ehsan Masood

South Africa reforms university funding

[CAPE TOWN] The South African government has proposed a number of mechanisms for bringing the growth in student numbers into line with the country's manpower requirements. They include completely restructuring the government's formula for calculating university subsidies.

In particular, the government wants to separate the state funding of universities' research and teaching components. Research funds will be earmarked, as will funds for student financial support and those for redressing past imbalances between institutions.

The government's plans, contained in a draft white paper (policy document) published last month, also suggest that funding for undergraduate teaching should be allocated on the basis of fixed subsidies per degree place in each discipline, varying according to the costs of training a student in that discipline.

But it is unclear whether places will be allocated at faculty level or at the level of the particular programmes set to replace the current system of major subjects.

The state will set a national total for the number of subsidized places in each discipline, and individual institutions will negotiate their share of places for a three-year cycle through a Council on Higher Education (CHE). Research funds will be linked to the allocation of postgraduate places, and will be commensurate with training costs.

These funds will be allocated on the basis of existing research capacity and on the need for institutional development. But it is unclear how these two criteria will be assessed.

The new proposals contrast sharply with the present funding system, set up in 1982, which allocates funding using a formula incorporating student enrolment and success rates, but with different weighting for arts- and science-based courses, and for undergraduate and postgraduate students.

The current formula also includes a measure for research output. This will not feature in the new formula, but will be taken into account when student places are reallocated at the end of each three-year cycle.

With no limits on student numbers in the various disciplines, the present system has led to undirected growth, particularly in the arts and social sciences. The government acknowledges the resulting imbalance in the qualifications of graduates, but indicates that, although future growth will attempt to increase enrolment in science-based programmes, cuts are not envisaged in other disciplines, as an overall growth in student numbers is anticipated.

The draft white paper does not provide for the incorporation of colleges of education, nursing and agriculture into universities and technikons, as suggested last year by the

National Commission on Higher Education (see *Nature* 384, 13-14; 1996), although it does propose that they are administered under one system.

Universities are relieved that they will be allowed to retain their own admissions criteria, and not be required to work through a central admissions office. Hugh Amoore, registrar at the University of Cape Town, describes the draft white paper as a "great improvement" on the green paper (consultation document). In particular, he welcomes the concept of a three-year funding cycle, as that will help planning.

But Amoore and others warn that the success of the changes is predicated on competence at ministerial, civil servant and CHE level. They express concern, for example, that the senior post responsible for higher education in the Department of Education is vacant.

A draft Higher Education Bill, providing for the establishment of the CHE, has also been published, and should be put to parliament during this session. Institutions have until 23 May to respond to both, and it is hoped that the council will be appointed by October.

Universities are likely still to be funded according to the old formula in 1998, although this may be combined with the distribution of some earmarked funds. The new system is destined to be in operation from 2000.

Meanwhile, the continued reduction in South Africa's university subsidy under the old formula, announced in March, has led to further student unrest. Four universities, including the Universities of the North and Fort Hare closed down, with students protesting at a decision to exclude those who were unable to pay fees.

Michael Cherry

Researchers pin hopes on international sources

[CAPE TOWN] AS access of university researchers in South Africa to government funds declines, many are hoping that their salvation lies in international sources of funding.

This month, the European Union (EU) agreement on scientific and technological cooperation with South Africa, which was signed last December, takes effect, and several United Nations agencies have recently established offices in South Africa and are assisting with research development.

The agreement with Europe enables South Africa to participate in the union's fourth Framework research programme, where it is classified as a developing country. Europe provides 50 per cent of the costs of joint research programmes by groups with participants from two of its member states and at least one developing country.

One of the first South Africans to receive EU funds, Karen Esler of the botany department at the University of Stellenbosch, says that this will provide "a tremendous boost" to her research on arid ecosystems.

The United Nations Educational, Scientific and Cultural Organization (Unesco) last year established an office in Pretoria to serve southern Africa. According to its science and technology adviser, Benjamin Ntim, it is particularly interested in developing a programme to address the problem of inadequate laboratory equipment in schools, and it may enlist the help of other international agencies.

South African researchers are now eligible to apply for funds from Unesco's regular programme, in any field of science, or for specially earmarked funds for research on biotechnology, man and the biosphere, and hydrology. As a result of South Africa joining the organization's Intergovernmental Oceanographic Commission, its researchers are now eligible for sponsorship to attend its meetings and training workshops.

The World Bank has also recently opened an office in Pretoria, and is encouraging research projects that coincide with its mission of "a sustained attack on

poverty". The bank's deputy representative, Junaid Ahmad, emphasizes the need to engage local researchers in programmes focusing on economic issues and behavioural changes.

The Global Environment Facility, administered by the bank and the United Nations Development Programme, is considering several applications from South African research groups.

Last month the Benefit project was launched (Benguela Environment Fisheries Interaction Training), with substantial aid from Norway and Germany. This is concerned with developing an understanding of the Benguela fisheries on the west coast of South Africa, Angola and Namibia, and of procedures for stock assessment.

Among the benefits for scientists from these countries will be access provided by the European partners to ships' time, which is expensive. "Nothing much has happened yet, but at last things are starting to happen," says John Field, president of South Africa's Scientific Committee on Oceanic Research.

M.C.

Research councils identify gaps in Europe's science

[MUNICH] Science in Europe is relatively weak compared to other countries — particularly the United States — in astroparticle physics, organic synthesis, fuel cell chemistry and the study of coastal ecosystems. These are some of the gaps that have been identified in the first survey analysing the strengths and weaknesses of European research, which was carried out under the auspices of the heads of European research councils and published last week.

The survey also confirms some familiar concerns. These include the lack of mobility of researchers, insufficient interdisciplinary research and a poor record of technology transfer coupled with lack of access to venture capital. It places particular stress on the declining stability of career structures for young scientists.

One of the scientific gaps identified by the survey was in classical optics, where Europe has previously played a leading role. The survey reported that Europe had lost its competitive edge to the United States in image processing and image recognition, and the storage of optical information. Another gap is in tropospheric chemistry, where the survey says technological advances in laser techniques and spectroscopy have not been applied.

The survey was carried out by Eurohorcs, an informal group of research council heads, at the request of the European Science and Technology Assembly (ESTA), the European Commission's advisory body (see *Nature* 383, 659; 1997). It is intended to help ESTA to identify weaknesses that could harm the competitive potential of Europe.

But the speed of the exercise — the survey was completed within two months — and the resulting variation in the quality of input has led some science policy experts to question its value.

The survey results are being circulated to professional organizations for comment, and will be discussed at the next ESTA meeting in July. Jan Borgman, the assembly's chairman, says it will use these comments to help gauge how much authority to give each of the reports.

Some of the scientific gaps identified in the survey will then be presented to the commission "to see how it responds". Borgman hopes that this can be done in time to influence the contents of the commission's fifth Framework research programme, which is due to be launched next year.

Despite the survey's weaknesses, Borgman says he is pleased with "this first attempt of the scientific community to design its own future".

Alison Abbott

Budget mismatch worries Australian scientists

[CANBERRA] Research leaders in Australia are expressing concern about an apparent mismatch between government claims that its budget, released last week, includes a growth in science funding, and official background documents showing a decline in real terms.

Joe Baker, president of the Federation of Australian Scientific and Technological Societies, says it was "absolutely amazing" that in his budget speech, Peter Costello, the Treasurer, had made "no reference to education, science, technology or innovation".

The speech contrasted sharply with a public commitment made three months earlier by the Prime Minister, John Howard, that the "central" role of science and technology "will never be ignored so far as my government is concerned".

In a statement issued on budget night, the minister for science and technology, Peter McGauran, said the government intended to increase direct Commonwealth spending for science and innovation by A\$43 million (US\$33 million) in 1997–98, to a total of A\$3.55 billion (see *Nature* 387, 222; 1997).

In particular, he said that the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia's chief science agency, would receive a five per cent increase in funding to A\$473 million.

But the minister's annual Science and Technology Statement, and his department's Portfolio Statement, both released later, reveal that when the sums are converted to constant dollars, there will be a decrease of A\$38 million — or 1.3 per cent — in the year beginning 1 July.

Even more alarming to researchers has

been an unannounced recalculation of the cuts from the 1996–97 budget. These were predicted at the time to represent 5.1 per cent in real terms, but the actual effect has been nearly twice as much — 9.6 per cent — the largest cut since statistics began.

The apparent overall increase for next year came largely from an extra A\$33.7 million added to the budget for CSIRO, and A\$6.7 million added for the Australian Nuclear Science and Technology Organisation (ANSTO). But detailed analysis shows that there will actually be no increase in operating funds for either agency in the coming year.

In CSIRO's case, the apparent funding increase results from a delay in returning to the government funds obtained from asset sales that the research organization was required to make in last year's budget.

CSIRO will also feel the effect of a A\$10-million cut over two years in the Cooperative Research Centres (CRCs) programme, in which it is the major partner. It must also achieve an 'efficiency dividend' of one per cent on its non-research and development activities, amounting to a A\$1.25-million cut.

When cuts announced to CRCs were not as great as had been rumoured, McGauran claimed that "there were champagne corks popping all around the CRCs". But a sense of celebration was missing from CRC directors, who expect that the number of centres will be reduced, and have warned that the cuts will "take away all the government's unallocated funds in future years".

The increase for ANSTO is entirely for the reprocessing of spent fuel rods at Dounreay in Scotland.

Peter Pockley

Germany plans 'unique' centre for Bonn

[MUNICH] Plans for a major new interdisciplinary research centre in Bonn were approved last week by Germany's Science Council, the Wissenschaftsrat. The new centre will be unique among federally supported research institutes in having the freedom to select and change its research areas at short notice.

The Centre of Advanced European Studies and Research (CAESAR) will begin work in temporary accommodation next year. It has been set up with part of a DM1.6-billion (US\$950-million) federal grant intended to develop the Bonn region as a major science area, in compensation to the city for the German government's move to Berlin (see *Nature* 374, 109; 1995).

CAESAR will be funded from a DM750-

million capital foundation. Most of this comes from the federal grant, which will be made available in instalments until 2004, and DM190 million of this will be spent on a new building and equipment. Interest from the remaining sum will support an international staff of around 150 scientists, all of whom will be hired on short-term contracts, with a maximum stay of ten years. An additional 150 scientists funded from other sources may also be accommodated.

The scientists will work in 60 small groups on "technologies of the twenty-first century", such as nanoscience, biotechnology and informatics.

Each group will work closely with industry and other international research institutions.

Quirin Schiermeier

Access law threat to medical merger

[SAN FRANCISCO] A shadow has been cast over the planned merger between the medical facilities of the University of California at San Francisco (UCSF) and Stanford University because of public concern about the transparency of decision-making.

Last week, the California state senate approved two bills that would require the two medical schools to comply with the state's laws on open records, open meetings and conflict of interest, making all information on the use of public funds available to the public.

In the past, Stanford has said it will pull out of the merger if such a law is passed, although there are signs that the two universities may be prepared to compromise with the senate's concerns. These will now be debated in the state assembly, which is considering two parallel bills.

The universities have agreed to merge their four hospitals, faculty practice plans and clinics into a private non-profit entity, UCSF-Stanford Health Care (USHC). The aim is to enhance the survival prospects of the medical facilities in today's 'managed care' marketplace. But there has been growing concern among UCSF faculty members, and others, about the "atmosphere of secrecy" surrounding the new corporation.

State senator John Burton (Democrat, San Francisco), who drafted the bill, said he felt California taxpayers had a right to information on the new corporation's operations, as \$300 million in public assets would be transferred to USHC. Burton believes that Stanford's threat to back out of the deal is a bluff, because the merger would benefit the university considerably.

Bruce Wintroub, professor of dermatology at UCSF and chief medical officer of the merged company, says the new organization would be a public benefit institution with a public mission. But he argues that it would be hard to compete with other private companies if its plans and records were open for public scrutiny. "We're dealing here in a world of facts, not in a world of wishes," he told a recent meeting of the UCSF academic senate.

At the senate's meeting last week, members said they feared a lack of public accountability could set the stage for a shift to for-profit values and abandonment of both institutions' teaching, research and care-giving priorities. "If the merger goes forward, we become just another healthcare business," said Vishu Lingappa, professor of physiology and medicine.

The senate, made up mainly of tenured, tenure-track and some clinical faculty members, voted 2-to-1 for further investigation of the merger. Many members at the meeting pushed for a vote on a statement supporting or opposing the administration's plans.

A task force set up by the senate to examine research and teaching implications of the

merger expressed concerns over the lack of detail in the agreement about promotion and hiring procedures. It claimed that there was also a lack of information about funding for research and clinical teaching, access to patient populations for research and a commitment to open research and faculty independence.

"These concerns have to be nailed down up front, even if it makes us a little less economically competitive," said Svein Oie, professor of pharmacy and chair of the task force.

Lee Goldman, chair of the department of medicine and interim faculty adviser to the USHC board, told faculty members that many of their concerns had been considered during the merger talks and that the merger would be likely to benefit clinical research. "A core statement in the charter is to fulfil the academic goals of the two institutions," he said. "Either can pull out if unhappy."

But Kenneth Melmon, professor of medicine and molecular pharmacology, and chair of a parallel task force at Stanford, says that none of the new company's decisions shows an understanding that the two institutions distinguish themselves from other health-care providers through their work in basic biology, clinical practice and teaching. Melmon says that faculty members at both schools are collaborating to ensure that principles reflecting a commitment to teaching and research are included in the merger agreement.

Among the areas that need attention are major differences in clinical faculty size, pay,

house staff activities and clinical emphasis, says Melmon. UCSF's clinical faculty is four times larger than Stanford's, even though it serves a similar patient population in terms of numbers, acuity and clinical visits.

Stanford has been far more successful than UCSF in technology transfer to commercial enterprises. On the other hand, UCSF spent \$66 million in unreimbursed costs for indigent care in 1995, while Stanford spent just \$22.2 million on its unreimbursed adult medical services.

Although it is planned to consolidate specialty units between the two schools, both have strong departments in cancer management, transplantation and other areas that would not easily be absorbed by the other.

According to Robert Wachter, associate chair of the UCSF Department of Medicine, groups from both schools have been meeting to work out quality improvement and practice guidelines that would benefit both schools.

Opponents of the merger say it is unnecessary for the public University of California to provide \$300 million in assets to a merged entity with Stanford when UCSF's Medical Center achieved a \$21 million profit in 1995-96.

But supporters say that the schools made modest operating gains only by significantly reducing departmental budgets. Economies of scale and the elimination of duplication will put both schools in a stronger financial position, they argue.

Sally Lehrman

Reborn, Captain Cook's ship sails home

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

[LONDON] A full-scale replica of Captain Cook's HMS *Endeavour*, seen here arriving in its 'spiritual home' of Whitby, is spending seven months touring 15 ports around the United Kingdom before departing for the United States. The ship, which was constructed using original eighteenth-century plans and documents from the National Maritime Museum, was launched

in Western Australia in 1993 by the HM Bark *Endeavour* Foundation. Richard Ormond, the director of the National Maritime Museum, says that he hopes *Endeavour's* tour around the United Kingdom will "give the opportunity for people to experience the conditions under which Cook and his men made their breakthrough voyages of discovery".

UK research council plans shake-up of biomedical awards

[LONDON] Britain's Medical Research Council (MRC) is finalizing details of a major shake-up in the way that it distributes research grants. In future, requests for programme grants will be considered only if they form part of a group submission. The council is also said to be planning to abandon individual 'project' grants.

Although no details of the proposed changes have been publicly released, it is believed that they are intended to streamline the administration of grants and provide a greater focus to the roughly £300 million (US\$480 million) a year of research activities that the MRC supports in universities and in its own research units.

Some researchers have expressed concern that the requirement that grant proposals be submitted as part of a bid from a consortium containing a minimum number — possibly three — of separate UK-based research groups could reduce their flexibility. There is a worry that high-performing groups might find themselves required to team up with less productive research teams.

But the MRC is said to feel that the new structures are an acceptable compromise between traditional investigator-initiated research projects and excessive central direction. The MRC, like other research councils, has been under strong pressure from the government to cut administrative costs. Details of the new arrangements are due to be released on 30 May.

Bomb pioneer warns of 'pure fusion' threat

[WASHINGTON] Hans Bethe, who was head of theoretical physics in the Manhattan Project, which developed the atomic bomb in the Second World War, has written to President Bill Clinton calling for a ban on research — including computational experiments "or even creative thought" — that could lead to new types of nuclear weapons. Bethe says various research approaches being pursued in the United States and in Russia could lead to "pure fusion" weapons, which would produce fusion without need for the "primary" fission stage which triggers existing hydrogen bombs.

The proliferation of such weapons would be difficult or impossible to control because their manufacture would not require the expensive fissile isotopes used in current devices. The 90-year-old physicist says that no research directed at new weapons is included in the stockpile stewardship programme at the US nuclear weapons laboratories. But he calls for a declaration banning it in the run-up to consideration of

the Comprehensive Test Ban Treaty by the US Senate, "to discipline the bureaucracy, and to reassure the world" of America's intentions.

Ho joins Montagnier in AIDS centres project

[WASHINGTON] The medical entrepreneur who recently lured Luc Montagnier, the co-discoverer of HIV, from the Pasteur Institute in Paris to New York city now plans to enlist David Ho, a leading AIDS scientist at the Aaron Diamond AIDS Research Center in New York, in a project to establish a chain of for-profit AIDS research and treatment centres in that city.

Bernard Salick, who will take on the project through a new company he has founded, Bentley Health Care, told *The New York Times* that his aim is to make New York "the center of research and the center of treatment" for AIDS. He said his enterprise has "very significant" backing from large drug companies, insurers and hospital chains in the United States and Europe.

Ho, he said, will be his chief adviser on the project. He added that he is also determined to enlist the help of Robert Gallo, the other HIV co-discoverer and a rival to Montagnier. Salick brought Montagnier to New York by endowing a \$4.5-million chair at Queens College, where Montagnier will run the Center for Molecular and Cellular Biology.

Baltimore to leave MIT to lead Caltech

[WASHINGTON] David Baltimore, the biologist and Nobel prizewinner, is to leave the Massachusetts Institute of Technology to become president of the California Institute of Technology (Caltech) in Pasadena. Baltimore has spent more than ten years clearing his name in an epic case of alleged scientific misconduct by one of his laboratory staff, which in 1991 forced his resignation as president of Rockefeller University in New York. He will take over at Caltech in the autumn. He says that he will continue to head a six-month-old project at the National Institutes of Health to revive the government's AIDS vaccine research (see page 323).

Electrical fire at CERN holds up experiments

[LONDON] The two largest accelerators at the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland, have been brought to a halt following an electrical fire in a power supply last week. Burning insulation gave rise to hydrochloric acid fumes, requiring a major cleaning operation to prevent corrosion of the unit, which feeds power to the Super Proton Synchrotron.

Apart from its own research capabilities,

the synchrotron is used to feed particles to CERN's most powerful machine, the Large Electron-Positron collider, which was due to be started up following an upgrade of its maximum collision energy to 184 GeV. Research programmes at the synchrotron and collider are likely to be delayed by two to three weeks.

US technology advocate heads back home

[WASHINGTON] Mary Good, the under-secretary for technology at the US Commerce Department and an energetic advocate of technology programmes in the Clinton administration, will leave her post on 2 June and return home to Little Rock, Arkansas, a spokeswoman confirmed. William Daley, the new commerce secretary, is not expected to name Good's successor at least until next month.

Good was tipped to succeed John Gibbons as science adviser to President Bill Clinton, and her departure follows reports that that position is to be filled by John Deutch of the Massachusetts Institute of Technology (see *Nature* 387, 112; 1997).

'Poached' eggs reveal dinosaur's identity

[LONDON] These eggs, laid by the dinosaur *Oviraptor* about 80 million years ago in what is now Mongolia, form part of a touring exhibition 'Dinosaurs of the Gobi Desert' which is on show at the Natural History Museum in London until 31 August.

Back in the 1920s, Henry Fairfield Osborn of the American Museum of Natural History (AMNH) thought that humanity had evolved in Asia. He sent the adventurer Roy Chapman Andrews (the model for Indiana Jones) to the Gobi to find out. No human bones were forthcoming. Instead, Andrews made the first discoveries of dinosaur eggs, and assumed that they had been laid by the herbivore *Protoceratops*. Nearby skeletons of the rarer *Oviraptor* suggested thievery, hence its name. This story was turned on its head after the AMNH returned to Mongolia in the 1990s and found new specimens. Embryos preserved in supposedly *Protoceratops* eggs belonged to *Oviraptor* — a case of poacher turning gamekeeper.

Campuses ring to a stormy clash over truth and reason

Temperatures remain high in US faculties as what started as a relatively esoteric debate on the nature of scientific knowledge continues to reverberate. Whatever their differences, both sides agree that important issues are at stake — including the role of universities themselves.

Universities across the United States are slowly awakening to a rumbling debate about the nature of scientific practice and knowledge, about who is qualified to pronounce on either — and about the motivation of those who do so.

You can call the argument the Science Wars, as some sociologists (and journalists) like to do. Or you can frame it in terms of the “flight from science and reason” that some scientists claim they can detect as the century draws to a close.

You can see it as a fundamental clash in perspectives between the sciences and the humanities — or as a petty squabble between two groups of academics, whose musings have little relevance to the outside world. But for some of those directly involved, particularly in disciplines such as the history and the sociology of science, the stakes are becoming very high indeed, with careers as well as academic credibility on the line (see page 332).

At the same time, for many of those who have started to follow the debate closely, the bitterness of some of the discussion has been counterbalanced by the sweet discovery that there is intellectual life on the other side of the still-yawning chasm between the two cultures, the sciences and the humanities, which appear to have grown no closer to each other since C. P. Snow identified the problem of their separateness 40 years ago.

Snow found his division between science and the arts and classics. Now, the most notable divide lies between science and the social sciences, whose size and importance in the academic community have exploded since Snow's assessment in the late 1950s.

“A lot of the shouting that's gone on has been most unfortunate,” says Trevor Pinch, of the science studies department at Cornell University, New York, of the recent debate, in which his own work has been one of the main targets of criticism. “But I see this as a great opportunity to address issues that need to be addressed and haven't been.”

Pinch and Harry Collins of the University of Southamp-

ton in the United Kingdom are the co-authors of *The Golem: What Everyone Should Know about Science* (Cambridge University Press, 1993), an influential text on the sociology of science. The book sets out to explain how science works in practice, and to discuss how much faith the public should place in it.

A ‘constructive dialogue’

The Golem has stirred considerable controversy among scientists, especially physicists. For example, the two authors have engaged in a lengthy and lively exchange of views with physicists in the pages of *Physics Today*, the journal of the American Physical Society.

Pinch sees one of the main issues at stake as being whether non-scientists, such as himself and others associated with the so-called ‘Edinburgh school’ (see page 333), can be accepted as making a valid intellectual contribution to contemporary scholarship in their attempts to study science and pose questions about how it works.

He believes he has achieved a constructive dialogue with scientists critical of sociological studies of science, such as Kurt Gottfried and Kenneth G. Wilson (see *Nature* 386, 545–547; 1997), and David Mermin, a crystallographer at Cornell University, New York, and *Physics Today* columnist.

Gottfried and Mermin are due to join Pinch and Collins at a meeting in July in Southampton to discuss their differences.

The physicist

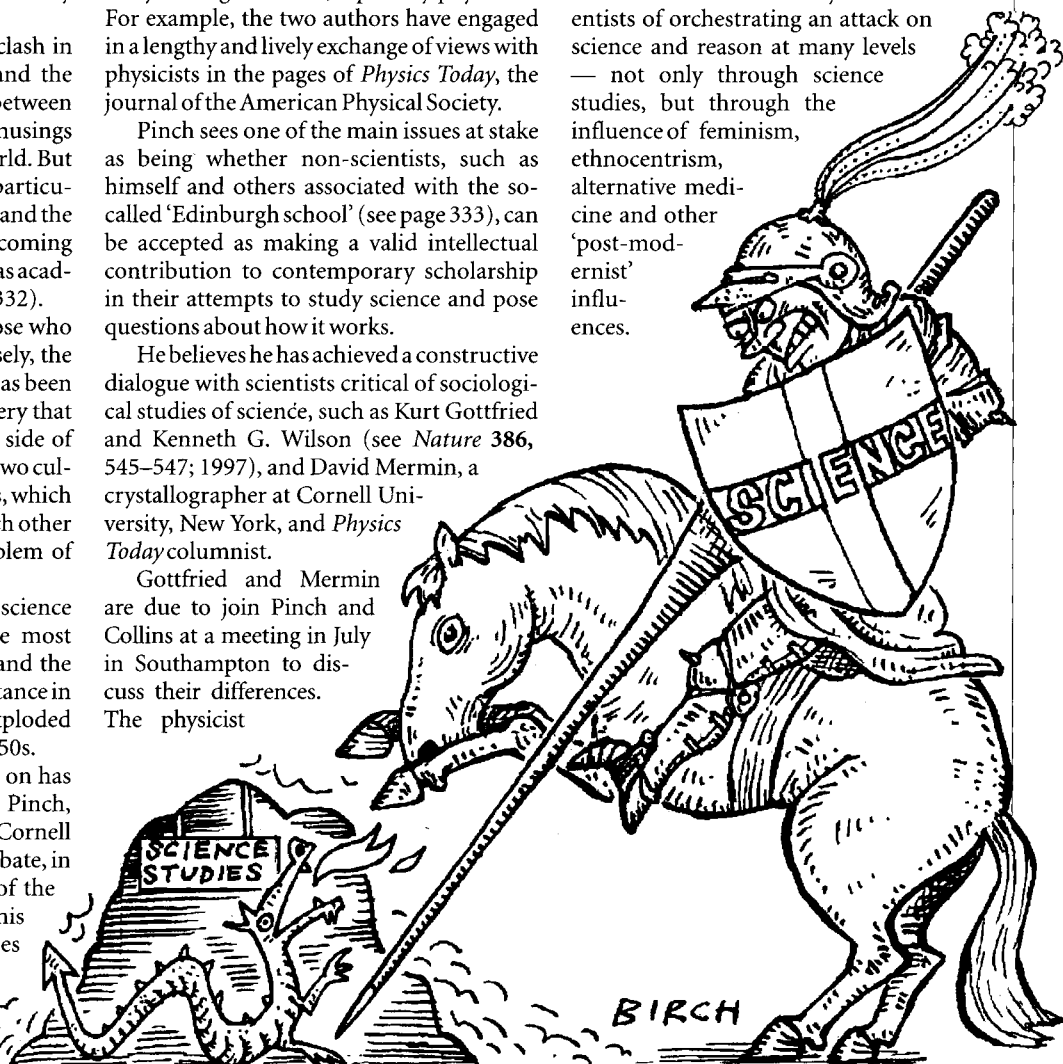
Steven Weinberg, a prominent critic of science studies, will also attend.

The serious and measured tone of the debate in *Physics Today* and elsewhere has raised hopes in the ‘science studies’ community that it may, in fact, represent the coming of age of their discipline. “The whole timbre of the discussion has become useful and productive,” says Collins.

But there is little prospect of much mutual respect or understanding between social scientists and the more outspoken group of scientists who kicked off the current debate three years ago, and whose fiery rhetoric has kept it on the boil since.

Two prominent members of this latter group are Paul Gross, a former director of the Marine Biological Laboratory at Woods Hole, Massachusetts, and Paul Levitt, professor of mathematics at Rutgers University in New Jersey. The two teamed up to publish a book, *Higher Superstition: The Academic Left and its Quarrels with Science* (Johns Hopkins University Press, Baltimore, 1994), which set the debate in motion.

Gross and Levitt are uncompromising with critics of science. They accuse social scientists of orchestrating an attack on science and reason at many levels — not only through science studies, but through the influence of feminism, ethnocentrism, alternative medicine and other ‘post-modernist’ influences.



On other pages | University fall-out: p332 | Edinburgh school: p333 | Public involvement: p335

"We believe that there is in the West, among professors and others who are paid, in principle, to think and teach, a new and systematic flight from science and reason," Gross declared at a meeting in 1995 at the New York Academy of Sciences. He added for good measure that this flight had "brought with it a truculent defence in the name of 'democracy' of New Age and traditional sophistry and charlatanism" (see *Nature* 375, 439; 1997).

Such rhetoric may not have done much to bridge the divide between the two cultures. But it certainly convinced some scientists that the best way to defend science was to attack its critics. One such was Alan Sokal, a physicist at New York University. Inspired, he says, by Gross and Levitt, Sokal raised the ante of the debate with a now-celebrated hoax, in which he planted a ludicrous article about quantum theory in *Social Text*, a sociology journal.

Defending the hoax later in the *New York Review of Books*, Weinberg said Sokal had

performed "a great service" by raising the profile of the debate. "We will need to confirm and strengthen the vision of a rationally understandable world if we are to protect ourselves from the irrational tendencies that still beset humanity," he thundered. But the Sokal hoax has been decried by many social scientists as going beyond the bounds of fair play. "I don't think discussion is furthered by poking fun at people," says one sociologist.

Collins makes a broader comment about the beginnings of the debate. "In the early days, it was as if we were entering this native village, and all the guard dogs came out and started biting our ankles," he says. "It was a mistake to



Sokal: 'two cultures' gap still remains.

try to have a discussion with the guard dogs."

But Sokal's activities have undoubtedly helped to broaden the issue, which has even found its way into the columns of *Newsweek* magazine. Now, at meetings across the United States, an increasing number of academics are joining in. Earlier this year at the University of Kansas, Lawrence, for example, more than 200 staff and students showed up to hear Sokal defend his unconventional tactic.

"I didn't have to invent anything" to produce the hoax, Sokal explains. "I simply quoted the post-modernist masters and showered them with praise." He admits that the success of the hoax proves nothing, except that one journal should more carefully review what it publishes. But Sokal has used the notoriety of the hoax to build a broader case against what he brands "the sloppy thinking and glib relativism that have become prevalent in many parts of science studies".

One of the sociologists attacked by Sokal and Gross, Steve Fuller, professor of sociology at the University of Durham in the United Kingdom, went to the Kansas meeting to defend science studies. "I won't deny that some people are in this to bash science — but only a very few," says Fuller.

Fuller argues that the idea of 'progress' in science and technology is a passing fad, which arose with the Enlightenment and is now diminishing as society realises that science and technology will not solve the problems that people used to hope that they would.

Converging on truth

This is one of the points at which dialogue between science and its critics breaks down. Most scientists do indeed believe that they are making absolute progress towards understanding nature. They think that the scientific method renders their knowledge quite distinct from other bodies of knowledge — such as the study of history, for example — where absolute truth is elusive, and the characters of the people involved and their chosen ways of working clearly play a major role in the direction ideas take.

Adrian Melott, a physicist at the University of Kansas, points out that scientists see the most important form of progress as being that of "convergence on the truth". He told Fuller: "You seem to view that as a defunct concept. I don't buy that. It is clear that we understand natural reality better than our predecessors did."

Sokal joined in on the attack: "Do you think that physics has made any progress in the last 300 years?" Fuller seemed to stall. "From our standpoint, yes; but could you convince these people [of Galileo's era] that you had advanced on what they'd done?" For most of the scientists in the room, the answer was obviously "yes".

But if the case for scientific progress remains strong, its presentation — through Gross, Levitt and Sokal in particular — has

Science studies braces for the fall-out

The Science Wars debate has sent a chill wind through a corner of the academic world that is more used to operating out of the limelight — namely the discipline known loosely as 'science and technology studies' or STS.

Critics of the area, which has struggled to establish its academic legitimacy ever since various universities decided in the late 1960s that scientists should be taught about the social implications of their work, argue that this wind should be seen as a breath of fresh air.

But many of those professionally involved prefer to talk in terms of witch-hunts, scapegoats and misunderstandings, arguing that they are being required to pay the price of a few youthful indiscretions, and are being unfairly blamed for social attitudes towards science whose origins lie elsewhere.

So far, the fall-out has been mostly anecdotal and focused on individuals. It is being widely claimed, for example, that the black-balling of a prominent historian of science, Norton Wise, from his appointment to the faculty of the Institute for Advanced Study at Princeton University, has been linked to his involvement in the so-called 'Sokal affair' (see page 325).

But there is little doubt that those responsible for teaching courses in the field feel that the tone of the debate has put them and many of their colleagues on the defensive. "The whole thing is creating an intense nervousness in our community," says Steven Shapin, a member of the original 'Edinburgh school' (see opposite) who now teaches the history and sociology of science at the University of California, San Diego.

Others claim more direct antipathy to their efforts. Tom Gieryn, for example, of the department of sociology at Indiana

University, said that the American Chemical Society, as sponsor of a controversial exhibit on Science in American Life at the Smithsonian Institution in Washington, had tried "to remove every mention of social science from the exhibition", for which he had been a consultant.

Gieryn was speaking at a meeting organized two weeks ago by the department of science and technology studies at Cornell University, the only US university to offer a full undergraduate degree in science and technology studies. Sheila Jasanoff, the chair of the department, points out that, even if the Science Wars debate is not having a direct impact at the institutional level, the fall-out from it may already be influencing funding decisions.

"When money is tight, you tend to cut out the things that are considered non-essential," says Jasanoff. "The curriculum in academic life is becoming more market driven than 'intellect' driven. The sociology of scientific knowledge is an easy thing to target, and appears to stand for a lot of things that some people say are going wrong in society."

The demand for such courses is still strong among undergraduates. Protests from students were a key factor that helped persuade Stanford University last year to make available an extra tenured position, to avoid having to close its interdisciplinary STS programme (see *Nature* 383, 563; 1997).

Jasanoff also claims that growing links are being forged between individuals trained in STS-related topics — for example in studies of the public perception of risk, or of the status of scientific evidence and expertise — and those engaged in confronting such issues in a practical way in political and legal arenas. **D. D.**

taken a curious form. For a start, as one scientist at the Kansas meeting observed, it is odd that, if it is science that is under attack, nearly all the aggressive language in the debate comes from the scientists' side.

This has led some observers to link the debate with cuts in federal funding for science, and especially for physics. "The sciences have different goals from the humanities," says Barry Shank of the American studies department at Kansas University. "That is not the problem. The problem is that funds are dwindling, and that has produced tension."

Noticing that much of the criticism of science studies has come from physicists, some have drawn a very sociological conclusion — that a weakened physics community has been letting off some of its frustration by lashing out at its perceived enemies in the social sciences.

Universities at stake

But such arguments irritate Gross, a biologist, who generously suggests that physicists and mathematicians are "maybe more intellectually adventurous" than life scientists. "Is it reasonable to think that people like Weinberg started to make a fuss about this because he was getting his funding cut? It's ridiculous," says Gross.

Gross attributes the timing of the argument to an entirely different factor — the collapse of radicalism on university campuses, and the associated "deep disappointment of intellectuals in the politics they have practised". He believes that a tide of relativism, which holds that no idea is more 'truthful' than any other, is filling the gap left behind. Although their arguments can often sound conservative, Gross says he used to be a leftist, and both Levitt and Sokal still are.

So why should busy researchers concern themselves with what might appear to be a minor tiff on the fragmented intellectual left? The view that unites all the protagonists in this debate is one unfashionable in modern America — namely, that intellectuals matter.

"University life has a powerful and direct influence on the way people think," says Gross. For him, after 40 years in the academic world, most recently as provost of the University of Virginia, the danger lies in "rapid growth of socio-political nonsense" on campuses which, he fears, will permeate the rest of society.

Those that Gross criticizes in the social sciences also want to ensure that intellectuals develop an accurate and truthful understanding of the world to convey to the rest of society. But for these social scientists, it is the way that science is practised and portrayed, as much as the results of experiments, that are in need of exposition.

In a world that depends as never before on science, the conflict strikes at the heart of arguments about the role of scholarship in modern society. Battles of this type are seldom short, and often bloody. No one pretends this one is over yet.

Colin Macilwain

Champions or challengers of the cause of science?

One of the ironies of the Science Wars debate is that the academic unit from which many of the now-disputed ideas emerged was set up partly to bridge the very gap — that between the famous 'two cultures' characterized by C. P. Snow — that these ideas seem to have exacerbated.

At the time of its creation in the mid-1960s, the Science Studies Unit at the University of Edinburgh was intended to help attract students into science by broadening their education. Few of those involved expected it would be criticized as it is now for undermining faith in science itself.

What has become known in many circles as the 'Edinburgh school' originated in the work of three individuals who were brought together on the staff of the unit in the early 1970s: Barry Barnes, a sociologist, David Bloor, a philosopher of science, and Steven Shapin, a historian.

It was these three in particular who developed the so-called 'strong programme' as a method of studying science that, although since complemented by other contrasting methodologies, has come to play a key role in defining the sociology of scientific knowledge.

According to David Edge, a former astronomer who became the unit's first director, the unit's goal in launching this initiative was self-consciously to extend the scientific method.

Bloor says that the details of the 'strong programme' emerged from the efforts by the three researchers to forge a joint approach to particular issues relating to how science is carried out in practice, while at the same time retaining their respective disciplinary identities.

"It was from the three of us talking together that a common idiom emerged as a precondition of being able to communicate,"

says Bloor. "Through homing in on a common perception of a set of problems, we realised that what we were doing was properly called the sociology of scientific knowledge."

The 'strong programme' seeks to study the social dynamics of scientific debates without assuming or drawing conclusions about the correctness of one side or the other of an argument — just as anthropologists, from whom many of its techniques are taken, might study a social group's beliefs without making judgements about their correctness.

The 'symmetric' approach

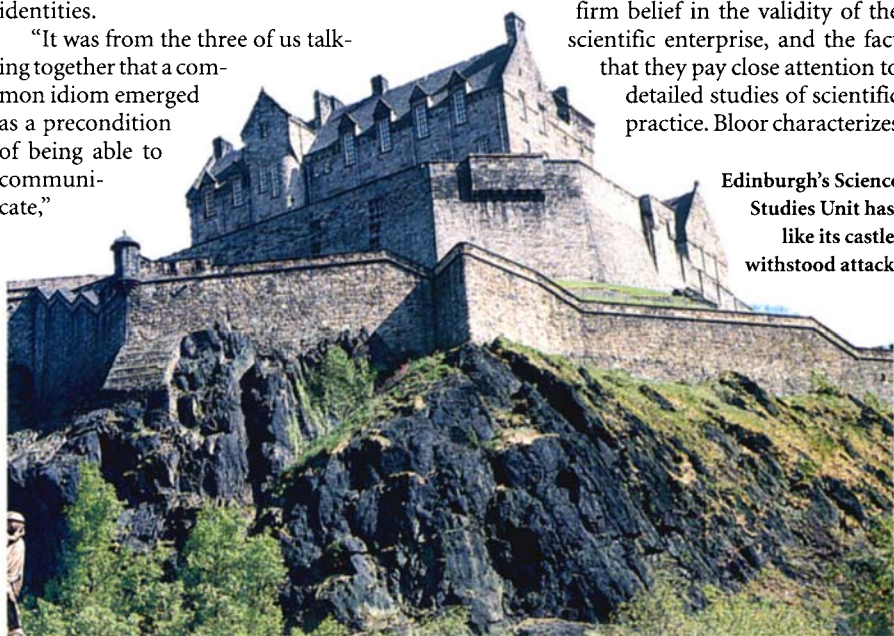
"We see the sociology of scientific knowledge as part of the project of science itself, an attempt to understand science in the idiom of science," say Barnes, Bloor and John Henry in their book *Scientific Knowledge: A Sociological Analysis*¹, which provides an overview of the 'Edinburgh' approach.

The decision to hold back from making judgements about rival hypotheses, referred to as the 'symmetry principle', has opened the Edinburgh researchers and their followers to the charge that they hold modern science to be merely a belief system, with no substance based in the reality of the world as revealed through experiments.

Physicists Kurt Gottfried and Kenneth G. Wilson, for example, in a recent article², concentrated their critique of the Edinburgh school on what they claimed to be its contention that "scientific knowledge is only a communal belief system with a dubious grip on reality".

But Bloor and his colleagues vigorously deny this criticism, pointing to their firm belief in the validity of the scientific enterprise, and the fact that they pay close attention to detailed studies of scientific practice. Bloor characterizes

Edinburgh's Science Studies Unit has, like its castle, withstood attack.



WORLD PICTURES

his goal, encapsulated in the strong programme, as being "to understand the nature of knowledge in a scientific manner". He adds: "I have been continually struck by the fact that philosophers of science were not actually working like scientists."

Bloor says that he, Barnes and Shapin share a common curiosity about science. The problem, he suggests, is that "we do not want a philosophical celebration of science, but a matter-of-fact analysis of it". Bloor argues that it is this approach that appears to have upset people, admitting that he is "shaken and shocked" by the reaction.

Shapin, who now teaches the history and sociology of science at the University of California in San Diego, describes his own role at Edinburgh as being "to give some of the basic ideas we were working on an empirical grip". Now widely known for several key historical studies, including a recent reinterpretation of the Scientific Revolution of the seventeenth century, Shapin reacts sharply against the argument that the work of the group ignores the importance of experiments. "We have a great respect for what scientists actually do," he says.

He suggests that part of the reason for the hostility towards the Edinburgh school lies

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Popper: too 'rational'
for science's good?

"If our fundamental philosophical position was anti-anything, it was anti-rationalist philosophy, not anti-science," says Shapin. "I find it very odd that the scientists who are upset are those that are hitching their wagon to impoverished and unsustainable philosophies. Why do they see criticism of Popper as criticism of science?"

Honouring science

Indeed, Shapin suggests that, rather than pursuing an 'anti-science' agenda, the sociologists of science are "doing scientists an honour by showing that science is a much richer, more complex and more heroic activity" than some rationalist explanations imply.

Shapin's arguments on this score are shared

in the way that it deliberately set out to provide a critique of a rationalist philosophy of science — the idea that scientific practice follows a set of established and accepted rules — as expressed by philosophers such as Karl Popper and Imre Lakatos.

by Harry Collins, now professor of sociology at the University of Southampton, author of an influential study of scientific practice³, and co-author with Trevor Pinch of the controversial volume *The Golem*⁴. Collins, who started his research in the sociology of science at the University of Bath in the early 1970s, is not a product of the Edinburgh school, although his name is often linked with it by others.

Indeed, there are some clear tensions between the sociologists themselves. Members of the Edinburgh group openly express their disagreement with what they claim to be the 'idealism' in the work of Collins, who appears to argue that scientific ideas are determined less by the behaviour of the material world than by that of the scientists who examine this behaviour.

Collins describes this as "a theoretical argument that has been going on for 25 years", but says that the differences separating him from the Edinburgh school are relatively unimportant compared to what they have in common, especially when it comes to detailed studies of scientific work.

Some argue that the Edinburgh school holds less influence today that its critics think, pointing out that its popularity was already being overtaken in the late 1980s by writers such as the French sociologist Bruno Latour, who rejects 'social constructivist' ideas and directly studied how scientists actually carry out research in the laboratory, not merely what is written about it.

"In the 1970s, the 'strong programme' was the intellectual spine against which any other school of thought in the sociology of scientific knowledge had to define itself, in the same way that Latour and the people around him became the centre of things in the 1980s," says one sociologist.

Other philosophers remain sceptical of the whole endeavour, arguing that supporters of the 'strong programme' — ironically in common with some of the philosophical rationalists they criticize — are wrong to ignore the important ways in which science, however chaotic an activity, does enable the emergence of basic truths about nature.

"The sociologists of science have done a lot of exciting and interesting empirical work," says David Papineau, professor of philosophy at Kings College, London. "But they are wrong in their view that this experimental approach shows that science is 'constructed', and that it does not get at something that we can call the truth." But that interpretation is challenged by members of the Edinburgh group, who claim to be looking at how truth is 'discovered' through being 'constructed'.

David Dickson

European sparks fail to ignite

Despite a few light skirmishes (see opposite), and the fact that the work of various European researchers has been dragged into the conflict in the United States, the Science Wars have so far failed to ignite into a major conflagration in Europe itself.

Earlier this year there were signs that it might be about to happen. The fuse appears to have been lit when articles stimulated by the 'Sokal hoax' appeared in newspapers such as *Le Monde* in France, Germany's *Die Zeit*, and Britain's *Times Literary Supplement* (see *Nature* 385, 381; 1997).

There have also been isolated instances in which aspects of the conflict have flared up in individual universities. One historian of science, for example, failed to be appointed to a chair at one European university after newspaper coverage about his possible appointment — although there is no clear evidence that his academic views were responsible for the decision.

But there has been little of the violent reaction to the sociology of scientific knowledge that has become a prominent feature of the US academic scene over the past couple of years. Nor do the growing number of groups that teach such subjects at universities feel themselves threatened in the same way.

Some claim that this is because the achievements of the area known as 'science and technology studies' (STS) in analysing the impact of social pressures on the content of science and technology are becoming

increasingly embedded into important aspects of government strategy.

"One of the success stories of STS in Europe is its significant contribution to the techniques of 'constructive technology assessment', which is becoming accepted in countries such as the Netherlands," says Wiebe Bijker of the University of Maastricht, who has already given seminars on 'social constructivism' to top government officials.

Bruno Latour, of the Ecole des Mines in Paris — one of the authors whose work has been both influential and frequently criticized in the United States (as well as in Europe) — suggests that support for European scientists is not as directly dependent on public opinion as it is in US politics, and is not therefore felt to be as threatened by public criticism.

"In many European countries, science has grown up embedded in [the machinery of] the state," says Latour. "This makes scientists feel more secure." Latour experienced US reaction to his ideas about how science works at first hand when he was turned down for a position at the Institute for Advanced Study at Princeton University several years ago.

He points out that in France virtually no natural scientists have taken part in debate about Sokal's article, and that its impact has been primarily among social scientists, some of whom agree strongly with Sokal on the limitations — and occasional absurdities — of post-modernist writers.

D. D.

1. Barnes, B., Bloor, D. & Henry, J. *Scientific Knowledge: A Sociological Analysis* (Athlone Press, London, 1996).

2. Gottfried, K. & Wilson, K. G. *Nature* 386, 545–547; 1997.

3. Collins, H. M. *Changing Order: Replication and Induction in Scientific Practice* (Routledge & Kegan Paul, London, 1992).

4. Collins, H. M. & Pinch, T. J. *The Golem: What Everyone Should Know about Science* (Cambridge Univ. Press, Cambridge, 1993).

Gunfire echoes in debates on public understanding

While US campuses, academic journals and newspapers remain the main theatres of action in the 'Science Wars', Britain has had its own 'science skirmish', with scientists and sociologists at loggerheads over the question of public involvement in science policy.

The two conflicts, while not identical, share many characteristics — and involve some of the same actors. Those who dismiss claims that scientific knowledge is essentially a social construct also tend to feel that setting priorities for research is essentially a task that should be left to scientific 'experts'.

In contrast, a broader view of science that challenges some of its claims to possess a unique insight into rational thought finds some of its strongest supporters among those who also argue that the public should contribute to the decision-making process.

The focus of the British debate has been on what is generically described as 'the public understanding of science' (broadly equivalent to the US concept of 'scientific literacy'). In 1985, the Royal Society published what some consider to be a seminal report exhorting scientists to communicate their work more often, with more enthusiasm, and more effectively to the public.

High profile science

The report led to the setting up of a committee on the public understanding of science (COPUS). And the following decade has seen a plethora of related activities, with more science reports in the media, courses in how to communicate science more effectively, and a national science week.

For some, such as Sir Walter Bodmer, principal of Hertford College, University of Oxford, and the author of the Royal Society study, such achievements represent a triumph that is too often underrated. The past decade, he believes, is an example of how "Britain leads the world", not only in science itself, but also in improving its public understanding.

But others, such as Brian Wynne, director of the Centre for the Study of Environmental Change at the University of Lancaster, are more circumspect. Wynne questions the extent to which more public exposure for science will contribute to an issue he believes to be equally important: fostering public trust in science, at a time when issues such as 'mad cow disease' appear to have dented the public's confidence in science.

Wynne believes that not only does the public have a right to know about publicly-funded science, but that they also have a right to influence decisions about science.

Such a notion is controversial. There is arguably no more hostile an opponent to



Bodmer: public must not control science.

increased public participation than Lewis Wolpert, professor of biology as applied to medicine at University College London, the present chairman of COPUS, and a sharp critic of the views of sociologists of science.

Wolpert says he is all in favour of improving public access to science. He says he is keen to encourage an understanding of the process as well as the history of science, and to involve the public in debates about the implications of science. But, he says, the shaping of science policy should be left to the "experts". He emphasizes, however, that this is his personal view and not necessarily that of COPUS.

"Science policy is a very technical issue," says Wolpert. One or two lay people sitting on advisory committees "will do no harm". But, he adds, "you can't have the public deciding whether to spend money on particle physics or astronomy. They're simply not qualified to make those sorts of decisions. Frankly, I don't think that the public cares."

Wolpert's views are shared to an extent by Bodmer, who says he is in favour of public involvement in science, "but not as a route to controlling scientific, particularly basic, research". Bodmer says it is a contradiction to involve the public in setting priorities for basic research, which is by definition open-ended and curiosity driven.

But Bodmer, who recently retired as director-general of the Imperial Cancer Research Fund, does not doubt the need for public involvement in issues such as medical and defence research, where the science has clearer implications for public policy. "My whole job at ICRF was to get the public to support us; otherwise there would have been no cancer research," he says.

Such views are widely held. Angela Wilkinson, head of public understanding of science at the Shell oil company, believes that the public care little about the detail of sci-

ence policy, but are primarily concerned that they should be able to trust any scientific information they are provided with.

The Brent Spar issue, in which public pressure led Shell to abandon its plans to sink the oil storage facility in the Atlantic, was "basically about people telling the government they did not like the idea of decisions being stitched up between industry and government behind closed doors," says Wilkinson. "This has never been the case. But we decided that if the government was not going to open up, we certainly would." Shell has since opened up decision-making about the future of the Spar to extensive public consultation.

Role for the public

Signs are, however, beginning to emerge of a change in thinking on the composition of advisory committees, in particular to dispel the idea that public involvement amounts to tokenism. Philip James, director of the Rowett Research Institute in Aberdeen, and author of a report commissioned by the Labour party on the setting up of a food standards agency, has gone as far as recommending that public and consumer interests should dominate the commission that manages the agency. In addition, James has suggested that advisory committees attached to the agency should meet in public.

If taken on board, the James recommendations are expected to open up another issue that some believe strikes at the core of the debate on public involvement in science policy. That question relates to defining the groups or type of people that would qualify for the title 'public'. In other words, should the word be applied to informed but disinterested members of the community at large, to members of advocacy organizations, or simply to individuals chosen at random, who may have very little knowledge of the issues?

Julie Shepherd, senior public affairs officer at the Consumers' Association, is convinced that public representatives need to be fully informed about issues they are asked to deal with. "There is this idea that the 'public' member of an advisory committee should be some Joe Bloggs, or a well turned-out lady from the shires," she says. "But we need consumer advocates who are equipped to challenge and take part in debate." **Ehsan Masood**



The one that got away: conventional interpretations of scientific 'evidence' are no longer adequate to contain science-based controversies, such as Greenpeace's protests against the sinking of Brent Spar.

The last natural nuclear fission reactor

Sir— The last known natural fission reactor on Earth is likely to be mined this year. Because these natural reactors are unique, at least one should be preserved for present and future research programmes.

The uranium deposits of Gabon host high-grade ores in which nuclear fission reactions occurred 2 billion years ago. These reactors have been studied since their discovery in 1972 because they represent a unique opportunity to describe and understand the behaviour of actinides and fission products in a geological system. The results of these studies have been used to establish analogies between the behaviour of geological systems and materials in natural reactors, and the materials and processes that can lead to the containment or migration of fission products and actinides in nuclear waste repositories (Final report EUR 16857: Oklo-Natural Analogue, Office for Official Publications of the European Communities, 1996). The natural reactors also provide unique information on the processes by which natural systems can reach criticality. There are no other known deposits of this type.

All the reactors except one are located in the most important uranium deposit of Gabon's Franceville basin, at Oklo. Fourteen reactors have been discovered in this deposit, all of which are now partially or completely mined. This deposit will be completely mined out soon, in 1998.

Future work on these reactors will therefore have to rely on previously collected samples, many of which are poorly documented and are out of their geological context. It will therefore not be

possible to study the behaviour and migration of fission products and actinides in the near-field and far-field environments of these reactors.

Such work is still possible, however, in a reactor located in the very small uranium deposit of Bangombé 30 km from Oklo. This reactor has not been mined. It is only 12 metres deep and has therefore been affected by near-surface weathering, but initial results show that the reactor is well preserved and that the effects of weathering on the distribution of actinides and fission products can be discerned. A European research programme (Oklo-Natural Analogue, Phase II) has been established for a coordinated and interdisciplinary study of this reactor in order to complete hydrogeochemical and geochemical modelling of the geological processes that affect radionuclide migration (for example, high-temperature hydrothermal alteration, low-temperature meteoric alteration, migration and sorption along fracture networks, and the mineralogy and geochemistry of the radionuclide-bearing phases).

The probability of other reactors on the Earth is very small because of the rather special conditions required for nuclear fission reactions to occur:

(1) The uranium deposit hosting the reactor must be older than 1,800 Myr, in order to have an adequately high $^{235}\text{U}/^{238}\text{U}$ ratio (the half-life of uranium-235 is 700 million years, as compared with 4.5 billion years for uranium-238).

(2) The deposit must have a low content of elements such as boron and vanadium,

which are neutron absorbers that decrease the possibility of the fission chain reaction beginning and being sustained.

(3) In order to preserve the reactor deposits for a long period, the deposit must be located in a very stable basin, protected from geological events that could lead to the remobilization of the uranium involved in the nuclear process. To our knowledge, this situation is not known to have existed anywhere else on the planet. Thus there is every reason to expect that the Bangombé reactor is the very last on the Earth.

We propose that this unique, scientifically important deposit be preserved for present and future research. This deposit is no less unique, and certainly more irreplaceable, than the most valued specimens from the Moon and Mars. Yet the deposit consists of only 100–200 tonnes of commercially viable uranium, so there is no compelling economic need to mine it. Prudence and responsibility require that careful consideration be given to the preservation of this last natural reactor.

F. Gauthier-Lafaye

Centre de Géochimie de la Surface, CNRS,
1 rue Blessig,
67084 Strasbourg,
France
e-mail: gauthier@illite.u-strasbg.fr

This note is signed by the scientists of the Technical Committee of the European Program (Oklo-Natural Analogue, Phase II):

P.-L. Blanc, J. Bruno, F. Gauthier-Lafaye, L. Griffault, E. Ledoux, D. Louvat, V. Michaud, M. Montoto, V. Oversby, L. Perez del Villar, J. Smellie

Who cares what's new?

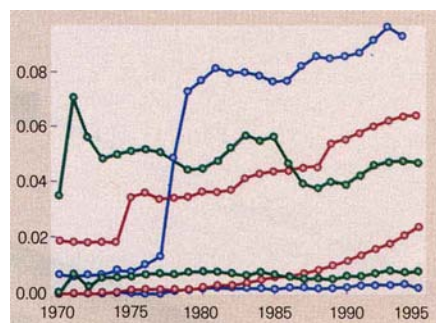
Sir— An exponential increase in the use of the word 'novel' by scientists to describe their work has been revealed in a statistical analysis of the Medline database by S. H. Friedman and J. O. M. Karlsson (*Nature* 385, 480; 1997). We have carried out a similar analysis of the MathSci and PsycLIT databases searching for titles and abstracts containing the words 'novel' or 'new'.

The figure plots frequency of the words 'novel' and 'new' as a proportion of total papers. The Medline data are also shown for comparison.

In contrast with the exponential increase seen in the Medline database, the use of the word 'novel' in the PsycLIT database has remained essentially constant over the past twenty years. There is only a slight increase in use seen in the MathSci database.

The use of the word 'novel' in the

PsycLIT and MathSci databases is well below that in the Medline database. By contrast the use of the word 'new' in the MathSci database showed a tenfold increase between 1975 and 1980.



Proportion of papers per year in the indicated databases that contain the words 'novel' (lower points) and 'new' (upper points) in the title or abstract. Blue denotes MathSci, red Medline and green PsycLIT.

There is also a significant increase in use of 'new' seen in the Medline database, but in the period 1980–94 this use has remained well below that in the MathSci database. The use of the word 'new' in the PsycLIT database has remained essentially constant over the past twenty years.

A possible explanation is that what is 'new' to the mathematical scientists is both 'new' and 'novel' to the biomedical scientists but neither here nor there to the psychologists and behavioural scientists.

Murray T. Batchelor

Department of Mathematics,
Australian National University,
Canberra ACT 0200, Australia

Bruce I. Henry

Simon D. Watt

Department of Applied Mathematics,
University of New South Wales
Sydney NSW 2052,
Australia

The end of the World is not nigh

Sir — My book *The End of the World: the Science and Ethics of Human Extinction*, reviewed by Freeman Dyson (*Nature* **380**, 296; 1996), joined Richard Gott (*Nature* **363**, 315; 1993) in developing a point originated by Brandon Carter. People have some grounds for distrusting any theory that makes them very unusually early among all humans who will ever have been born. Taking account of this, they can justifiably re-estimate the risk that humankind will soon be extinct. They can become less confident in Dyson's suggestion (*Reviews of Modern Physics* **51**, 246; 1979) that their descendants will continue onwards eternally.

Dyson calls it "typical" of the book to reason as follows: that if the number so far born is about a hundred billion then "our species has only a ten-per-cent chance of surviving as long as five hundred generations with our present size of population". In fact, however, the book follows Carter in arguing only for changes in the risk-estimates generated by examining various dangers. If Carter and I saw a gigantic asteroid rushing at us, or well funded plans for surviving asteroid

impacts, then that would affect our probabilistic calculations. What Dyson calls typical bears scant resemblance to anything we have said. Further, we are innocent of the blunder he describes. We avoid assuming simultaneously (1) that, as he puts it, "we know nothing of our place in the history of our species", and (2) that we know we are placed "among the first hundred billion humans".

Consider the following story, told in the book. If fully funded, an experiment was sure to run as follows. Initially, three people would each receive an emerald. Several centuries later, emeralds would be given to 5,000 people. You yourself get an emerald. Imagine you are ignorant of your place in the population history of emerald-getters, but are certain the experiment was fully funded.

You should bet you are in the group of 5,000. However, suppose you instead know you are in the group of three, much as we could know we are in the first hundred billion humans. This fortifies any suspicions you may have that the experiment was poorly funded, so that few emeralds will be distributed later. No blunder here! For you need not be assuming, first, that you know absolutely nothing of your place among emerald-getters, so as to be able to ask in all seriousness whether you are in the group of

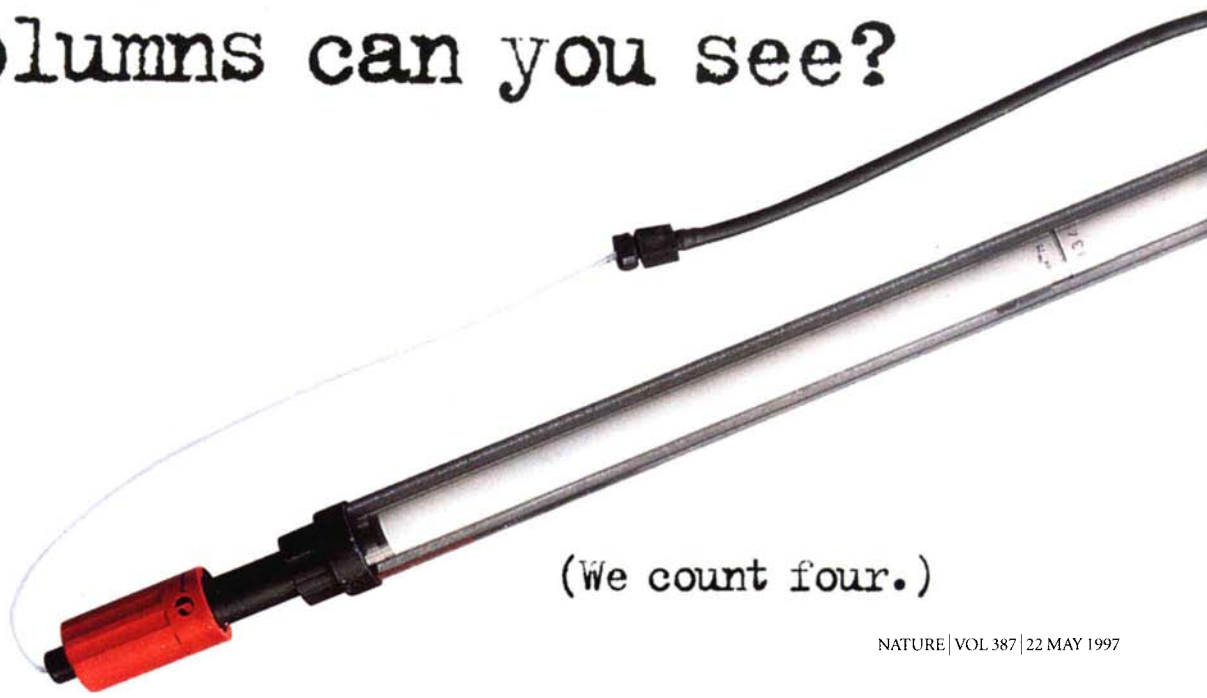
three, and second, inconsistently, that you know you are in it. Instead, certain you are among the three, but unsure whether there will be 5,000 emerald-getters later, you can be rightly taking account of how likely you as an emerald-getter had been to find yourself among the three, against the background of competing theories about the funding.

Had the funding been so poor that only three emeralds would ever be distributed, then the likelihood was unity; with full funding, it was roughly 0.0006. Taking account of such likelihoods is not claiming to be ignorant of what, in the very next breath, you say you know.

Take a simpler case. Two urns each contain three black balls. In one, there is nothing else; in the other, an additional 5,000 white. Having no idea of which is which, you choose an urn by tossing a coin. A ball is drawn, and it is black. You conclude that there is little chance that the urn is the one with the 5,000 white, for the likelihood of drawing a black ball from that urn had been so small. This reasoning does not involve declaring first that you are ignorant that a black ball has been drawn, and second that you know it. Asking what the likelihood would have been, on various hypotheses, that you would observe what you actually did observe, is fundamental to all science.

How many
gel filtration
columns can you see?

(We count four.)



Dyson adds that the book fails to "examine critically" the literature of the dangers confronting us. Can this be true? It evaluates lengthy arguments not only from doomsayers but also from their critics, concluding on page 146 that humankind's chances of avoiding extinction in the near future are "encouragingly high".

John Leslie

Department of Philosophy,
University of Guelph,
Ontario N1G 2W1,
Canada
e-mail: johnlesl@uoguelph.ca

There aren't plenty more fish in the sea

Sir— Six pages of Briefing on fisheries science (*Nature* 386, 105–110; 1997) ably demonstrates why the fisheries sector is in crisis. The reality is that if fishermen meet the market demand for fish, fish stocks will be completely fished out within a very short time.

There is only one answer to the fisheries crisis: we must farm more fish. Every fish the market receives from farming means that one less fish needs to be caught from the sea. The problem is that there is no universal species or farming technique that

will supply world needs. Appropriate requirements need to be identified on a local basis. Such limitations deter those looking at global solutions.

Unfortunately, this simple solution to the fisheries problem is unlikely to become a reality, for two reasons. The first is that traditional fisheries scientists have little knowledge of aquaculture and therefore do not perceive it to be a realistic solution. The second reason, and one that simply confirms the first, is that poor publicity from one or two examples of aquaculture where things have not gone quite right have damaged the potential for the aquaculture industry.

The example used in your Briefing is of shrimp farming, not fish production. There is no doubt that the commercial exploitation of farmed shrimp has led to problems. That is because previously extensive production units have simply been intensified beyond their capacity. While there is a clear focus on short-term rewards, there is also an increasing demand for these shrimp from the West, which has been met by this uncontrolled intensification.

Such over-exploitation of extensive farming techniques is the inevitable result of a fisheries sector that has marginalized aquaculture. This applies equally to Europe, where the European Commission has

decided that aquaculture development should be determined by individual commercial ventures rather than by defined long-term strategic policies. Aquaculture features only as a structural policy, not as part of any grand plan.

In Europe, while ministers argue about whether they should cut fishing by 20% or 30%, they are missing the opportunity to develop large-scale low-cost fish production. This is not comparable to the coastal pond systems used to produce shrimp in Asia, but instead is a form of offshore open-sea farming on a huge scale. Not only would this supply much of our fish requirement, but it would also create many needed jobs for skilled fish handlers.

Fisheries policies, as outlined in the Briefing, do not seem to have worked. Maybe it is now time to go back to the drawing-board and re-evaluate our fisheries needs. If we drop all our preconceived ideas and start again with a clean sheet, maybe we could guarantee the survival of our fish stocks. The alternative is that fish will disappear from both the seas and our diet.

M. R. Jaffa

Callander McDowell,
9 Haversham Court,
Middleton Road,
Manchester M8 4JY, UK

"Who would have thought a gel filtration column would make it so easy for me to get into *Nature*," says Stacy, a post-doc working in New York City.

That may look like just one column, but it's not. You see, that's Superdex™ inside that column. Superdex gel filtration technology is based on the best of the underlying technologies of Sephadex®, Superose® and Sepharose®—each already proven in thousands of published research findings. So, inside a Superdex column, you get the best technologies.

The best qualities of three great technologies in one: Superdex

Superdex delivers extremely rapid separations, incredibly steep selectivity curves, and truly minimal non-specific interactions. So it's ideal for purifying oligonucleotides, peptides, proteins, or other major biomolecules. The ingenious combination of the properties of dextran and agarose makes Superdex ten times faster than comparable gel filtration media. The proof of this statement is defined in the new Pharmacia Biotech "Gel Filtration: Practices and Principles"—a seventh edition handbook (that's right, edition number seven).

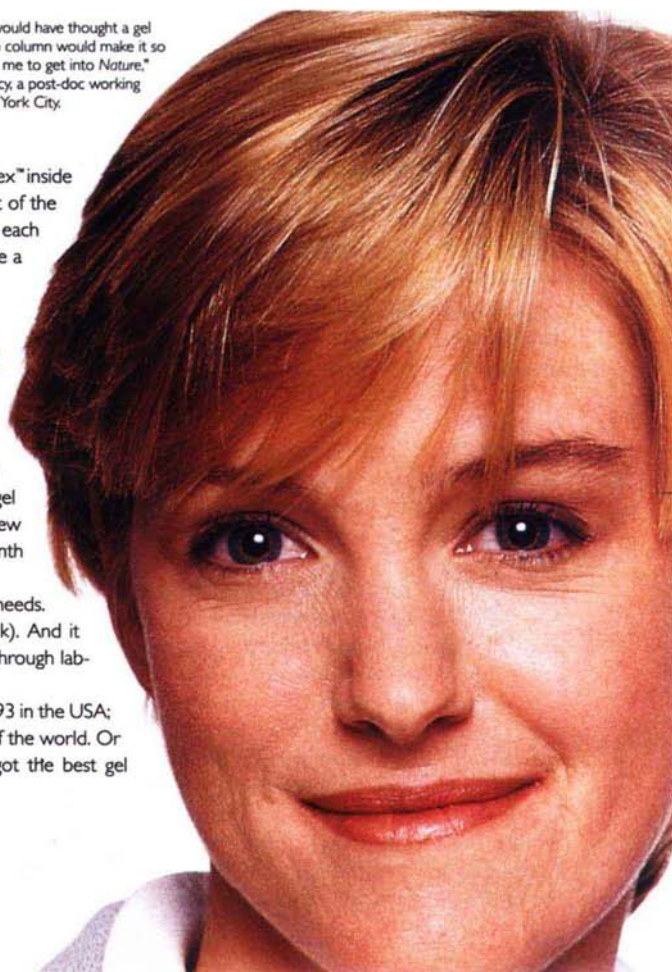
We're sure you'll find a Superdex selectivity to satisfy your needs. After all, it's available in three selectivity ranges (prep grade in bulk). And it comes in packs and columns varying from trial-size scouting packs, through lab-to process-scale columns—many of which arrive pre-packed.

Find out more about Superdex. Give us a call: 1 (800) 526-3593 in the USA; +81 3492 6949 in Japan; +46 18 16 50 11 in Europe and the rest of the world. Or visit us on the Internet: <http://www.biotech.pharmacia.se>. We've got the best gel filtration technology solution for you.



**Pharmacia
Biotech**

Uppsala, Sweden. (And the rest of the world)



Backing the wrong horse

Sir—As the author of the foreword to Stephen Budiansky's book *The Nature of Horses*, I was amazed by the inaccuracies in Marian Dawkins' review (*Nature* 386, 341–342; 1997).

The reviewer's first sentence is incorrect—nowhere on the book jacket or in any of the book's pages does the quoted blurb appear*. If the reviewer means some publicity the publisher sent with the review copy, then potential purchasers will never have seen it anyway. The reviewer then grumbles that the book doesn't live up to "its claim to cover everything one might want to know about horses". In fact, the subtitle is "Exploring Equine Evolution, Intelligence and Behavior", not "everything".

The reviewer also complains that the book "draws upon the work of other people". Most books do, but Budiansky is scrupulous in citing sources.

Franklin M. Loew

Medical Foods, Inc.,

201 Broadway, 5th Floor,

Cambridge, Massachusetts 02139, USA

*The offending phrase appeared in proofs of the book. Before Dr Dawkins' review appeared, *Nature* should have sent her a finished copy of the book. Editor, *Nature*

Space can wait

Sir—Yet another new (in effect) species has been discovered in the Annamite Mountains on the Laos/Vietnam border¹. It is going to cost millions, perhaps billions, of dollars to find out whether there might have been a bacterium on Mars in the distant past. It probably didn't take more than a few thousand dollars to rediscover *Sus bucculentus*, which is not a bacterium but a pig, a massive lifeform, as are a number of the other mammals still being discovered through the second half of this century².

Why do we spend so much to discover life on Mars and so little on discovering new life on our own Earth? There's another contrast too. It's not going to make the slightest difference to what we find on Mars whether we go now or in a hundred or even a thousand years' time. If the bacterium's traces are there now, they'll be there then. And the same goes for the search for new stars and galaxies, but more so: the Universe isn't going to look very different 10,000 years down the line from the way it does now. But one small part of it will be very different. If we don't find out now about present and new lifeforms on Earth, we will have lost the chance for ever: more and more species are 'dead as a dodo'. We can leave exploration of the stars and Mars to our great-grandchildren, our ten-greats

grandchildren. We cannot leave exploration of the Annamites, or any of the other threatened areas of Earth³, even to our children.

I am not anti-astronomy: Hale-Bopp is going to be a memory of a lifetime. But space can wait; Earth can't. I have a suggestion. For every new observatory, every new Hubble (every new repair of Hubble!), exactly matching funds from the same source should be automatically provided for the discovery and identification of new lifeforms on Earth.

Alexander H. Harcourt

Department of Anthropology,

University of California,

Davis, California 95616, USA

e-mail: ahharcourt@ucdavis.edu

1. Groves, C. P., Schaller, G. B., Amato, G. & Khounbolin, K. *Nature* 386, 335 (1997).

2. Pine, R. H. *Nature* 368, 593 (1994).

3. Morris, D. W. *Nature* 373, 25 (1995).

USGS managers should consult customers

Sir—The recent controversy about US Geological Survey Library cutbacks (*Nature* 386, 631; 1997) is just one of a series of highly questionable management decisions from the Geologic Division as they are incapable of finding a foothold in shifting political sands.

Over a period of two years, the Geologic Division, through buyouts and a reduction-in-force (RIF), lost more than one-third of its staff, including many key programmes and tremendous 'corporate' memory, experience and wisdom. These staff reductions, in effect, boosted their budget by more than \$50 million from a nearly constant congressional appropriation of \$250 million. How ironic that out of that surplus, a half-million dollars cannot be found to continue support of the library while millions are being spent to fight highly questionable RIF practices, including making women scientists redundant at a rate 70 per cent greater than their male counterparts, as well as all scientists with pending equal opportunity employment (EEO) complaints, and retaining all former managers (except the female who had a pending EEO complaint). This is a battle that will surely continue in the courts for many years to come even though settlements could be negotiated at a fraction of the continuing legal costs.

In large part, these controversies are generated because of the parochial attitude of the current Geologic Division management. While lip-service is given to serving customers, very little input is sought from those customers. The Geologic Division should realize that there are many companies and individuals, including those

made redundant, willing to provide input and help to re-establish its stature as a venerable and unbiased institution. The continued arrogant exclusion of intellectual diversity by the Geologic Division greatly hampers the organization's ability to survive in a fast-changing political environment.

G. Michael Reimer

122 S. Devinney Street,

Golden, Colorado 80401, USA

e-mail: mreimer@mines.edu

Public perception

Sir—A strongly worded resolution of the European Parliament recently established that procedures for marketing genetically modified organism (GMO) products should be revised, opposing the European Commission's rule that a transgenic maize could be marketed. This was underpinned by concern about safety for consumers and the environment, and that decision-making procedures did not reflect the opinion of the European member states, where 13 of 15 had opposed the procedure.

But how had the member states reached their decisions? In Britain, little systematic information exists about public attitudes to genetic engineering of foodstuffs. A recent report (*Uncertain Worlds*) from the University of Lancaster highlights ambivalence towards the use of GMOs in foodstuffs. Our recently completed study explores the prevalence of views among 386 sixth-formers—the coming generation of voters and purchasers. More than half (61%) were 'a bit worried' and a further 11% 'very worried' about genetic engineering. Some 59% agreed that it lengthened shelf-life; 58% thought it extended the growing season. Fewer saw advantages of taste (10%), economy (19%) or 'healthiness' (9%). Only 5% thought genetically engineered foodstuffs unsafe to eat, and 61% said they would eat such products. For 17%, genetically modified plants were an environmental risk. In several issues many students gave 'neutral' responses, so people may remain persuadable. In contrast, 86% thought genetically engineered foods should be labelled.

Such appreciation of lay people's views is essential if there is to be a claim to 'democracy'; views can be represented only if they are known.

Ruaraidh Hill

Martin Stanisstreet

Edward Boyes

Environmental Education Research Unit,

University of Liverpool,

Liverpool L69 3BX, UK

e-mail: martstan@liv.ac.uk

Helen O'Sullivan

Liverpool Hope University College,

Hope Park, Liverpool L16 9JD, UK

Nepotism and sexism in peer-review

In the first-ever analysis of peer-review scores for postdoctoral fellowship applications, the system is revealed as being riddled with prejudice. The policy of secrecy in evaluation must be abandoned.

**Christine Wennerås and
Agnes Wold**

Throughout the world, women leave their academic careers to a far greater extent than their male colleagues¹. In Sweden, for example, women are awarded 44 per cent of biomedical PhDs but hold a mere 25 per cent of the postdoctoral positions and only 7 per cent of professorial positions. It used to be thought that once there were enough entry-level female scientists, the male domination of the upper echelons of academic research would automatically diminish. But this has not happened in the biomedical field, where disproportionate numbers of men still hold higher academic positions, despite the significant numbers of women who have entered this research field since the 1970s.

Reasons for lack of success

Why do women face these difficulties? One view is that women tend to be less motivated and career-oriented than men, and therefore are not as assiduous in applying for positions and grants. Another is that women are less productive than men, and consequently their work has less scientific merit. Yet another is that women suffer discrimination due to gender. We decided to investigate whether the peer-review system of the Swedish Medical Research Council (MRC), one of the main funding agencies for biomedical research in Sweden, evaluates women and men on an equal basis. Our investigation was prompted by the fact that the success rate of female scientists applying for postdoctoral fellowships at the MRC during the 1990s has been less than half that of male applicants.

Our study strongly suggests that peer reviewers cannot judge scientific merit independent of gender. The peer reviewers overestimated male achievements and/or underestimated female performance, as shown by multiple-regression analyses of the relation

between defined parameters of scientific productivity and competence scores.

In the peer-review system of the Swedish MRC, each applicant submits a curriculum vitae, a bibliography and a research proposal. The application is reviewed by one of 11 evaluation committees, each covering a specified research field. The individual applicant is rated by the five reviewers of the committee to which he or she has been assigned. Each reviewer gives the applicant a score between 0 and 4 for the following three parameters: scientific competence; relevance of the research proposal; and the quality of the proposed methodology. The three scores given by each reviewer are then multiplied with one another to yield a product score that can vary between 0 and 64. Finally, the average of the five product scores an applicant has received is computed, yielding a final score that is the basis on which the applicants to each committee are ranked.

The MRC board, which includes the chairmen of the 11 committees, ultimately decides to whom the fellowships will be awarded. Usually each committee chooses between one and three of the top-ranked applicants. Of the 114 applicants for the 20 postdoctoral fellowships offered in 1995, there were 62 men and 52 women, with a mean age of 36 years, all of whom had received a PhD degree within the past five years. Most of the female applicants had basic degrees in science (62 per cent), and the rest had medical (27 per cent) or nursing (12 per cent) degrees; the corresponding figures for the male applicants were 38, 59 and 3 per cent.

Traditionally, peer-review scores are not made public, and indeed the MRC officials initially refused us access to the documents dealing with evaluation of the applicants. In Sweden, however, the Freedom of the Press Act grants individuals access to all documents held by state or municipal authorities. Only documents defined as secret by the Secrecy Act are exempt, for example those that may endanger the security of the state, foreign relations or citizens' personal integrity. Accordingly, we appealed against the refusal of the MRC to release the scores.

In 1995, the Administrative Court of Appeal judged the evaluation scores of the MRC to be official documents. Hence, to our knowledge, this is the first time that genuine peer-reviewer evaluation sheets concerning a large cohort of applicants has become available for scientific study.

We found that the MRC reviewers gave female applicants lower average scores than

male applicants on all three evaluation parameters: 0.25 fewer points for scientific competence (2.21 versus 2.46 points); 0.17 fewer points for quality of the proposed methodology (2.37 versus 2.54); and 0.13 fewer points for relevance of the research proposal (2.49 versus 2.62). Because these scores are multiplied with each other, female applicants received substantially lower final scores compared with male applicants (13.8 versus 17.0 points on average). That year, four women and 16 men were awarded postdoctoral fellowships.

As shown by these figures, the peer reviewers deemed women applicants to be particularly deficient in scientific competence. As it is generally regarded that this parameter is related to the number and quality of scientific publications²⁻⁵, it seemed reasonable to assume that women earned lower scores on this parameter than men because they were less productive. We explored this hypothesis by determining the scientific productivity of all 114 applicants and then comparing the peer-reviewer ratings of groups of male and female applicants with similar scientific productivity.

Productivity variables

We measured the scientific productivity of each applicant in six different ways. First, we determined the applicant's total number of original scientific publications, and second, the number of publications on which the applicant was first author. Both figures were taken from the applicant's bibliography, which we double-checked in the Medline database. (We call these measures 'total number of publications' and 'total number of first-author publications'.)

To take into account the fact that the prestige of biomedical journals varies widely, we constructed measures based on journals' impact factors. The impact factor of a scientific journal is listed in the independent Institute of Scientific Information's *Journal Citation Reports*, and describes the number of times an average paper published in a particular journal is cited during one year. Our third measure was to add together the impact factors of each of the journals in which the applicant's papers were published, generating the 'total impact measure' of the applicant's total number of publications.

Fourth, we generated the 'first-author impact measure' by adding together the impact factors of the journals in which the applicant's first-author papers appeared. The unit of measure for both total impact and first-author impact is 'impact points',

...the credibility of the academic system will be undermined in the eyes of the public if it does not allow a scientific evaluation of its own scientific evaluation system.

with one impact point equalling one paper published in a journal with an impact factor of 1.

Fifth, using the science citation database, we identified the number of times the applicant's scientific papers were cited during 1994, which yielded the measure 'total citations'. And sixth, we repeated this procedure for papers on which the applicant was first author, giving the measure 'first-author citations'.

Did men and women with equal scientific productivity receive the same competence rating by the MRC reviewers? No! As shown in Fig. 1 for the productivity variable 'total impact', the peer reviewers gave female applicants lower scores than male applicants who displayed the same level of scientific productivity. In fact, the most productive group of female applicants, containing those with 100 total impact points or more, was the only group of women judged to be as competent as men, although only as competent as the least productive group of male applicants (the one whose members had fewer than 20 total impact points).

Why women score low

Although the difference in scoring of male and female applicants of equal scientific productivity suggested that there was indeed discrimination against women researchers, factors other than the applicant's gender could, in principle, have been responsible for the low scores awarded to women. If, for example, women were mainly to conduct research in areas given low priority by the MRC, come from less-renowned universities, or have less collaboration with academic decision-makers, their lower scores could depend on such factors, rather than on their gender *per se*.

To determine the cause of women's lower scores, we performed a multiple-regression analysis, which reveals the factors that exert a primary influence on a certain outcome (for example competence scores) and the size of such an influence. Multiple regression permits the elimination of factors whose influence on a certain outcome merely reflects their dependence on other factors.

In the multiple-regression analysis, we assumed that the competence scores given to applicants are linearly related to their scientific productivity. We constructed six different multiple-regression models, one for each of the productivity variables outlined above. In each of these models, we determined the influence of the following factors on the competence scores: the applicant's gender; nationality (Swedish/non-Swedish); basic education (medical, science or nursing school); scientific field; university affiliation; the evaluation committee to which the applicant was assigned; whether the applicant had had postdoctoral experience abroad; whether a letter of recommendation accom-

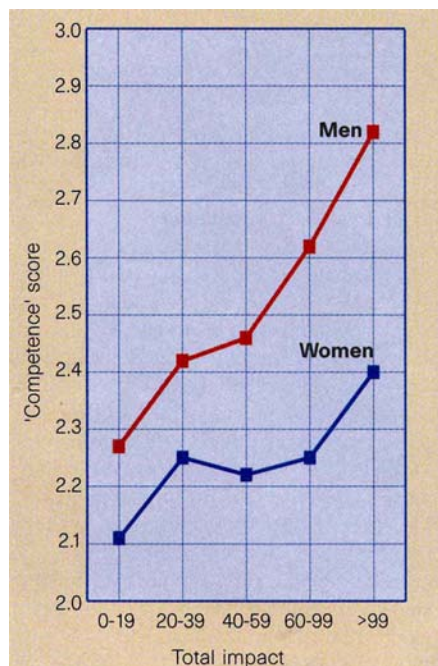


Figure 1 The mean competence score given to male (red squares) and female (blue squares) applicants by the MRC reviewers as a function of their scientific productivity, measured as total impact. One impact point equals one paper published in a journal with an impact factor of 1. (See text for further explanation.)

panied the application; and whether the applicant was affiliated with any of the members of the evaluation committee. The last piece of information is noted on the MRC evaluation protocols, in which case the reviewer in question is not allowed to participate in the scoring of that applicant. It was as frequent for female (12 per cent) as for male (13 per cent) applicants to be associated with a committee member.

The outcome of the regression analysis is shown in Table 1. Three out of the six productivity variables generated statistically significant models capable of predicting the competence scores the applicants were awarded: total impact, first-author impact and first-author citations. The model that provided the highest explanatory power was the one based on total impact ($r^2 = 0.47$). In all three models, we found two factors as well as scientific productivity that had a significant influence on competence scores: the gender of the applicant and the affiliation of the applicant with a committee member.

According to the multiple-regression model based on total impact, female applicants started from a basic competence level of 2.09 competence points (the intercept of the multiple regression curve) and were given an extra 0.0033 competence points by the reviewers for every impact point they had accumulated. Independent of scientific productivity, however, male applicants received an extra 0.21 points for competence. So, for a female scientist to be awarded the same com-

petence score as a male colleague, she needed to exceed his scientific productivity by 64 impact points (95 per cent confidence interval: 35–93 impact points).

This represents approximately three extra papers in *Nature* or *Science* (impact factors 25 and 22, respectively), or 20 extra papers in a journal with an impact factor of around 3, which would be an excellent specialist journal such as *Atherosclerosis*, *Gut*, *Infection and Immunity*, *Neuroscience* or *Radiology*. Considering that the mean total impact of this cohort of applicants was 40 points, a female applicant had to be 2.5 times more productive than the average male applicant to receive the same competence score as he ($((40 + 64)/40 = 2.6)$).

Friendship bonus

According to the same multiple-regression model, applicants who were affiliated with a committee member received competence scores 0.22 points higher than applicants of the same gender and scientific productivity who lacked such ties (Table 1). This 'affiliation bonus' was worth 67 impact points (confidence interval: 29 to 105 impact points). Hence, an applicant lacking personal ties with the reviewers needed to have 67 more impact points than an applicant of the same sex who was associated with one of the reviewers, to be perceived as equally competent. So, although MRC policy does not allow 'biased' reviewers to participate in the scoring of applicants they are associated with, this rule was insufficient, as the 'neutral' committee members compensated by raising their scores when judging applicants affiliated with one of their peers.

Because the affiliation bonus was of the same magnitude as the 'male gender' bonus, a woman applicant could make up for her gender (-0.21 competence points) by being affiliated with one of the reviewers ($+0.22$ competence points). On the other hand, a female (-0.21 competence points) lacking personal connections in the committee (-0.22 competence points) had to present an additional 131 impact points to the MRC reviewers to receive the same competence score as a male applicant affiliated with one of the reviewers.

Such a level of productivity was attained by only three of the 114 applicants, one male and two female. Hence, being of the female gender and lacking personal connections was a double handicap of such severity that it could hardly be compensated for by scientific productivity alone.

The two other regression models, based on first-author impact and first-author citations, yielded almost identical results to the first with regard to the effect of gender and affiliation (Table 1). This congruity was not a statistical artefact due to a high degree of interrelation between the three productivity variables, as the total impact and first-author

Table 1 Factors that significantly influenced peer reviewers' rating of scientific competence, according to three multiple regression models.

Multiple regression model based on:	Scientific productivity			Additional points given by the reviewers for the following factors			Size of the influence of the non-scientific factors in productivity equivalents		
	<i>r</i> ²	Intercept	Competence points per productivity unit	Male gender	Reviewer affiliation	Recommendation letter	Male gender	Reviewer affiliation	Unit of measure
Total impact	0.47	2.09	0.0033 <i><0.00005*</i>	0.21 <i><0.00005</i>	0.22 <i>0.0008</i>	0.10 <i>0.04</i>	64 (35-93)†	67 (29-105)	Impact points
First-author impact	0.44	2.13	0.0094 <i><0.0001</i>	0.24 <i><0.00005</i>	0.20 <i>0.005</i>	NS	25 (14-36)	21 (6-36)	Impact points
First-author citations	0.41	2.17	0.0054 <i>0.001</i>	0.23 <i><0.00005</i>	0.23 <i>0.001</i>	NS	42 (23-61)	42 (17-67)	Citations during 1994

* Italicized numbers indicate *P*-values for the variable in question.

† Numbers in parentheses indicate 95% confidence interval.

NS, not statistically significant, *P*-value > 0.05.

impact of the applicants were only moderately correlated ($r=0.63$), as were total impact and first-author citations ($r=0.62$). We therefore believe that male gender and reviewer affiliation were real determinants of scientific competence in the eyes of the MRC reviewers.

The applicant's nationality, education, field of research or postdoctoral experience did not influence competence scores in any of the models. A letter of recommendation had a positive effect on the competence score in the model based on total impact, but not in the two others (Table 1). By contrast, the evaluation committee that rated individual applicants did influence competence scores, as some committees were 'sterner' in their evaluation of competence than the rest (data not shown). However, an applicant who was assigned to a 'tough' committee had the same chance of being awarded a fellowship as other applicants, as fellowships were distributed based on the rank the applicant acquired within his or her committee and not on absolute score values.

Changing the system

The peer-review system, characterized as "the centrepiece of the modern scientific review process"⁶, has been criticized on many grounds, including poor inter-reviewer reliability² and because reviewers may favour projects confirming their own views⁷. Our study is the first analysis based on actual peer-reviewer scores and provides direct evidence that the peer-review system is subject to nepotism, as has already been suggested anecdotally⁸⁻¹⁰.

One might argue that young researchers affiliated with peer reviewers are part of a scientific elite that has received superior training and are therefore more competent than average applicants. Indeed, applicants with such ties had higher total impact levels on average than applicants without such connections (data not shown). Hence, applicants with personal alliances justly benefited from higher competence scores because of their higher scientific productivity. However, on top of that, they were given extra competence points not warranted by scientific productivity. We see no reason why an applicant who manages to produce research of

high quality despite not being affiliated with a prestigious research group should not be similarly rewarded.

Several studies have shown that both women and men rate the quality of men's work higher than that of women when they are aware of the sex of the person to be evaluated, but not when the same person's gender is unknown¹¹⁻¹³. It is somewhat surprising that the results of these studies have not discouraged the scientific community from relying on evaluation systems that are vulnerable to reviewer prejudice.

An interesting question that we could not address here is whether the harsher evaluation of female researchers was due to the paucity of women among the peer reviewers. The small number of women reviewers (5 out of 55) and their uneven distribution among the MRC's committees made a statistical analysis of their scoring behaviour impossible. However, a few studies have indicated that female evaluators may be more objective in assessing the achievement of women than their male counterparts¹⁴. Nevertheless, we are not confident that a simple increase in the percentage of women reviewers would solve the problem of gender-based discrimination.

If gender discrimination of the magnitude we have observed is operative in the peer-review systems of other research councils and grant-awarding organizations, and in countries other than Sweden, this could entirely account for the lower success rate of female as compared with male researchers in attaining high academic rank. The United Nations has recently named Sweden as the leading country in the world with respect to equal opportunities for men and women, so it is not too far-fetched to assume that gender-based discrimination may occur elsewhere. It is therefore essential that more studies such as ours are conducted in different countries and in different areas of scientific research.

An in-depth analysis of other peer-review systems can be achieved only if the policy of secrecy is abandoned. We could perform our study only because of the Swedish Freedom of the Press Act. It is often claimed that secrecy in scoring will protect reviewers from improper influences. But our results cast

doubt on these claims. It has also been suggested that the recruitment of peer reviewers of high quality would be impeded if reviewers were not granted anonymity. Such fears seem to be exaggerated because, although reviewer evaluation scores have been accessible to everyone in Sweden since the court ruling of 1995, there have been no large-scale defections of peer reviewers from the evaluation committees.

Most important, the credibility of the academic system will be undermined in the eyes of the public if it does not allow a scientific evaluation of its own scientific evaluation system. It is our firm belief that scientists are the most suited to evaluate research performance. One must recognize, however, that scientists are no less immune than other human beings to the effects of prejudice and comradeship. The development of peer-review systems with some built-in resistance to the weaknesses of human nature is therefore of high priority. If this is not done, a large pool of promising talent will be wasted. □

Christine Wennerås is in the Department of Medical Microbiology and Immunology and Agnes Wold is in the Department of Clinical Immunology at Göteborg University, Guldhedsgatan 10, S-413 46 Göteborg, Sweden (e-mail: agnes.wold@immuno.gu.se).

Acknowledgements. We thank documentalist Ann-Marie Holst for assistance, and Maria Wold-Troell, Svante Wold and Christer Andersson for statistical advice. The study was supported by a grant from the Swedish Ministry of Education.

- Widnall, S. E. *Science* **241**, 1740-1745 (1988).
- Cole, S., Cole, J. R. & Simon, G. A. *Science* **214**, 881-886 (1981).
- Long, J. S. *Social Forces* **71**, 159-178 (1992).
- Sonnert, G. *Social Stud. Sci.* **25**, 35-55 (1995).
- Sonnert, G. & Holton, G. *Am. Sci.* **84**, 63-71 (1996).
- Glantz, S. A. & Bero, L. A. *J. Am. Med. Assoc.* **272**, 114-116 (1994).
- Ernst, E., Resch, K. L. & Uher, E. M. *Ann. Intern. Med.* **116**, 958 (1992).
- Forsdyke, D. R. *FASEB J.* **7**, 619-621 (1993).
- Calza, L. & Gerbisa, S. *Nature* **374**, 492 (1995).
- Perez-Enciso, M. *Nature* **378**, 760 (1995).
- Goldberg, P. *Trans-Action* **5**, 28-30 (1968).
- Nieva, V. F. & Gutek, B. A. *Acad. Manag. Rev.* **5**, 267-276 (1980).
- O'Leary, V. E. & Wallston, B. S. *Rev. Pers. Soc. Psychol.* **2**, 9-43 (1982).
- Frieze, I. H. in *Women and Achievement: Social and Motivational Analyses* (eds Mednick, M. T., Tangri, S. S. & Hoffman, L. W.) 158-171 (Hemisphere, Washington DC, 1975).

Nature adds: This article was peer-reviewed by three males.

Hot-spotting in the Pacific

Seth Stein

Combining the geometry of plate tectonics, and the computational advantage of digital sea-floor topography, provides a new method for locating plumes rising from the Earth's deep mantle and for tracing the history of plate motions.

Geologists delight in contemplating maps of global topography, because the major features beautifully illustrate the geometry describing the motion of the great plates of Earth's lithosphere. The most obvious features are those showing 'relative' motions between neighbouring plates. At the long mid-ocean ridges, hot material wells up from the deeper mantle and cools to form plates, about 100 km thick, which then diverge. At deep-sea trenches, one plate descends under another and re-enters the deep mantle. Past plate motions are shown by the fit of continents, which were split when new ocean basins formed.

Another class of feature, linear chains of seamounts, indicate the 'absolute' motion of plates with respect to the deep mantle. These chains, such as the spectacular one extending from the Hawaiian Islands northwestward across the Pacific to the Emperor seamounts (Fig. 1), are thought to form over 'hotspots', where plumes of hot material well up from deep in the mantle^{1,2}. As a plate moves over the hotspot, new volcanic seamounts form (sometimes reaching the sea surface as islands), and are then carried away from the hotspot by the plate's motion. Provided plumes are fixed with respect to the deep mantle, or move more slowly with respect to it than plates, the seamount chains they produce have trends showing the direction of absolute plate motion, and a progression of volcanic ages indicating the rate of absolute motion.

Both relative and absolute plate motions are described by the simple geometry of the rotation of rigid plates on a sphere about a pole. Hence, a classic introductory text explains how profound insights about topography and other consequences of plate motions (earthquakes, mountain building, volcanism, and so on) can be had using "a little geometry, a piece of paper, a pair of scissors, and a logical mind"³. An elegant analysis by Wessel and Kroenke, described on page 365 of this issue⁴, again demonstrates the power of plate-motion geometry, but using satellite-derived gravity measurements and the computer rather than paper and scissors.

Their paper addresses the issue of using seamounts to locate hotspots and study

absolute plate motions. The challenge has been that the approach used so successfully for the Hawaiian–Emperor seamount chain has proved more difficult to apply elsewhere. Along this chain, almost one hundred seamounts and islands have been dated radiometrically, and show a progression from young ages near Hawaii to ages of 65 million years old in the North Pacific⁵. As a result, it has been possible to determine successive rotation poles (stage poles) describing the absolute motion of the Pacific plate⁶.

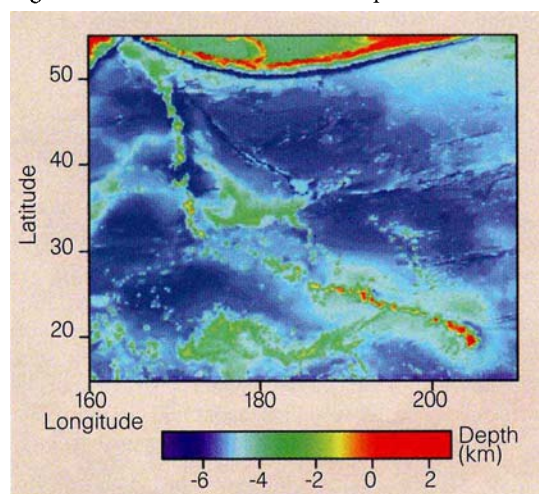


Figure 1 Hotspot track — bathymetry of the North Pacific, showing the seamount chain extending from the Hawaiian Islands (20° N, 205° E) northwestward across the Pacific to the Emperor seamounts. The seamounts and islands are the trace left on the Pacific plate by volcanism at the hotspot now under Hawaii. The bend in the chain near 32° N, 175° E is generally attributed to a change in absolute plate motion about 43 million years ago.

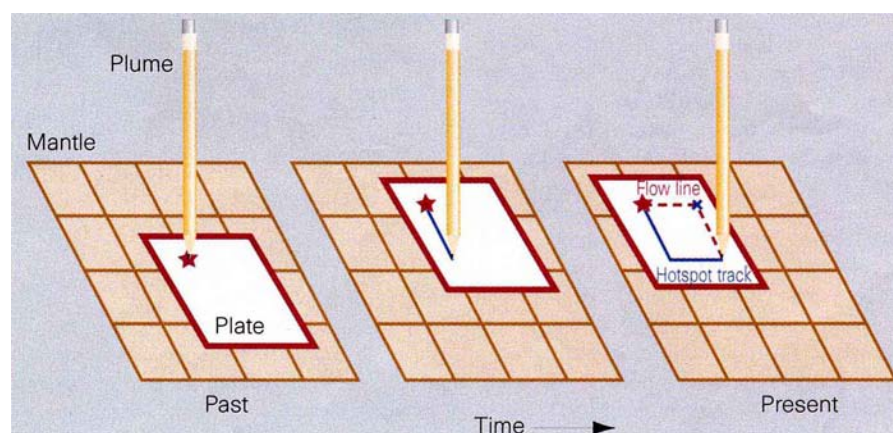


Figure 2 A simple experiment, demonstrating the principle used by Wessel and Kroenke⁴ to track the motion of a plate with respect to an upwelling plume (pencil) fixed in the deep mantle. The pencil line (blue) represents the volcanic seamount chain (hotspot track) produced as the plate moves over the mantle. This track differs from the flow line (red dashed line) showing the motion with respect to the mantle of the seamount (star) produced in the first panel. Because the flow line includes the hotspot location, the intersection of flow lines for different seamounts can be used to locate the hotspot that produced them.

For other chains, however, fewer sites have been dated, and the age progression is much less clear. Different workers have thus proposed either that some chains reflect the effects of several hotspots⁷ or that non-hotspot volcanism occurred synchronously along the chains⁸. The small number of dated seamounts also limits the accuracy of studies of past plate motions, and makes it difficult to determine whether hotspots move relative to one another⁹.

Wessel and Kroenke introduce a method for determining which seamounts were produced by hotspots, without knowing the age of the seamounts. The approach can be illustrated by a simple experiment shown in Fig. 2. Hold a pencil (representing a plume) fixed relative to a piece of paper with a grid (representing the mantle). Moving a piece of paper (representing a plate) between the two yields a pencil line (representing the seamount chain forming the hotspot track).

Consider a history in which, with respect to the mantle, the plate moved north for one time unit, and then west at the same speed for another time unit. The traditional approach

has been to infer the hotspot location by starting with a seamount (star) on the hotspot track, and 'backtracking' it through time by undoing the plate-motion history. This approach requires that the seamount age be known. In Fig. 2, for example, if the seamount is incorrectly assumed to be one (rather than two) time units old, backtracking undoes only the most recent plate motion, predicting that the hotspot that formed the seamount is at the site marked with a cross.

The alternative is to consider the path, or flow line, actually followed by the seamount over the deep mantle. Because the plate motion occurred in two stages, the backtracked path representing the chain of seamounts on the plate differs from the flow line. The two paths provide different ways of thinking about the motion between the seamount, which is fixed with respect to the plate, and the hotspot, which is assumed fixed with respect to the deep mantle. The flow line is the path the plate (and hence the seamount) followed with respect to the mantle (that is, the hotspot) once the seamount formed at the hotspot. The hotspot track is the path with respect to the plate (or seamount) of the hotspot since it formed the seamount. The flow line can be constructed without knowing the seamount's age, by assigning a range of ages to the seamount, from very young to older than the true age, and backtracking the seamount for each assumed age. So, in Fig. 2, the site with the cross found by backtracking with an assumed age of one time step lies on the flow line.

For the spherical Earth, the flow line and hotspot track reflect the same set of rotations about the stage poles, but in the opposite order. The two paths differ because finite rotations do not commute (as can be illustrated by performing successive rotations about a horizontal and vertical axis in different order on two books)³.

The flow line is less intuitive than the hotspot track, because it is not marked by a chain of seamounts, but has the property that it always passes through the hotspot location. So two seamounts formed at the same hotspot have flow lines that intersect there. In contrast, flow lines for seamounts not produced at hotspots generally do not intersect.

To apply the method, Wessel and Kroenke used 8,800 Pacific seamounts identified in the satellite-derived gravity field, which reflects sea-floor topography. Although few of the seamounts have known ages, and many were presumably not produced by hotspot volcanism, all can be used in the analysis. Seamounts were summed numerically along flow lines, giving a cumulative volcano amplitude (CVA) image whose maxima correspond to likely hotspot locations. The results illustrate several applications of the technique, which the authors

term 'hot-spotting'. By changing the assumed absolute motion of the Hawaiian hotspot for the past three million years from that previously assumed, the CVA maximum better matches the recent volcanism. This change also relocates the maximum corresponding to the Louisville hotspot in the southeast Pacific to a geologically plausible site where shallow earthquakes and volcanism may be occurring, better fits the observed trend of the Marquesas Islands, and predicts that the complex age pattern of volcanism in French Polynesia reflects the effects of several hotspots.

The hot-spotting technique should thus have wide application for determining which seamounts were produced at hotspots, locating these hotspots, and refining histories of absolute plate motions. It will also help address the long-standing but unresolved issue of whether hotspots are fixed with respect to one another or have some small but significant relative motion.

There may be several lessons here for Earth scientists. One is the value of serendipity: Wessel came to the flow-line concept by accidentally programming the rotations in the wrong order¹⁰. Another is the benefit of

fresh thinking: the idea of studying hotspots without considering seamount ages has been assessed as "it's phenomenal what can be done by throwing away half the data"¹⁰. Perhaps the deepest lesson is that the geometry of plate motions, a classic tool of post-1960 geology, not only remains powerful but can yield new insights. Geology departments might emulate Plato's Academy and inscribe over the door "Let no one ignorant of geometry enter". □

Seth Stein is in the Department of Geological Sciences, Northwestern University, Evanston, Illinois 60208, USA

(e-mail: seth@earth.nwu.edu).

1. Wilson, J. T. *Can. J. Phys.* **41**, 863–870 (1963).
2. Morgan, W. J. *Nature* **230**, 42–43 (1971).
3. Cox, A. & Hart, R. B. *Plate Tectonics: How it Works* (Blackwell Scientific, Oxford, 1986).
4. Wessel, P. & Kroenke, L. *Nature* **387**, 365–369 (1997).
5. Clague, D. A. & Dalrymple, G. B. in *Geology of North America: The Eastern Pacific Ocean and Hawaii* (eds Winterer, E. L., Hussong, D. M. & Decker, R. W.) 188–217 (Geol. Soc. Am., Boulder, CO, 1989).
6. Jarrard, R. D. & Clague, D. A. *Rev. Geophys. Space Phys.* **15**, 57–76 (1977).
7. Henderson, L. J. thesis, Northwestern Univ., Evanston, IL (1985).
8. Schlanger, S. O. *et al. J. Geophys. Res.* **89**, 11261–11272 (1984).
9. Duncan, R. A. & Richards, M. A. *Rev. Geophys.* **29**, 31–50 (1991).
10. Schneider, D. *Sci. Am.* **224**, 22–24 (April, 1997).

HIV–cell fusion

The viral mousetrap

James Binley and John P. Moore

AIDS researchers and those who are interested in how human immunodeficiency virus (HIV) fuses with cells will now have a spring in their step, thanks to two new papers by Weissenhorn *et al.*¹ (on page 426 of this issue) and Chan *et al.*² in *Cell*. The papers are conceptually similar

— each group has co-crystallized two large fragments from the ectodomain of the transmembrane glycoprotein gp41. The analyses reveal similarities between the gp41 core structure and the influenza-virus HA2 protein, which is the model on which viral fusion mechanisms are based^{3,4}. Although

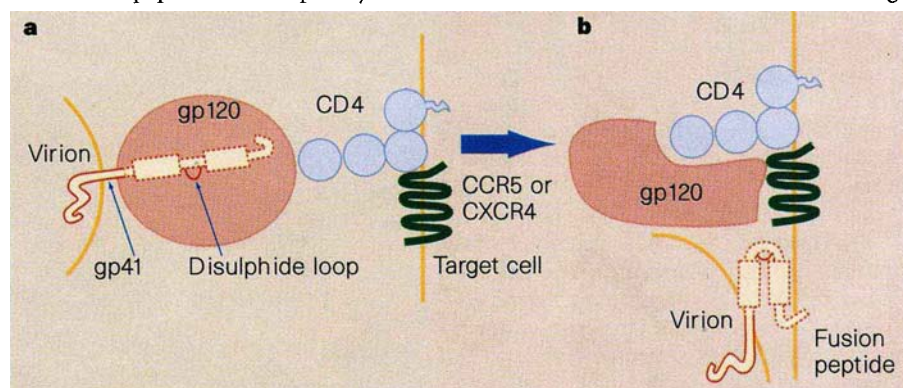


Figure 1 a, Human immunodeficiency virus (HIV)–cell fusion. The HIV gp160 precursor is post-translationally cleaved into the glycoproteins gp120 and gp41 which, although they occur as trimers, are depicted as monomers here. The pre-fusion conformation of gp41 is represented as an 'unsprung' mousetrap, although this native structure is, in fact, unknown. Gp120 binds CD4 on the target cell surface, then interacts with a co-receptor (CCR5 or CXCR4) near the cell membrane, triggering conformational changes in the envelope glycoproteins. These changes cause the gp41 'mousetrap' to snap closed, and the fusion peptide to penetrate the target cell. b, Weissenhorn *et al.*¹ and Chan *et al.*² have crystallized the ectodomain of the form of gp41 shown here. The two membrane-associative regions of gp41 (the transmembrane domain and the fusion peptide) are brought into close proximity when the 'mousetrap' is triggered, facilitating the coalescence of the viral and cell membranes and, ultimately, membrane fusion.

the influenza virus and HIV-1 are both RNA viruses, they belong to different families (the orthomyxoviridae and retroviridae, respectively), suggesting that the viral fusion process has been conserved — in its most essential features — throughout evolution.

Weissenhorn *et al.* co-crystallized fragments corresponding to residues 540–590 and 624–665 of the gp41-precursor, gp160; Chan *et al.* co-crystallized residues 541–576 and 623–656. These residues comprise about 54 per cent of the gp41 ectodomain — the internal segment of gp41 is probably not directly involved in fusion. In the complete gp41 ectodomain, the helical regions formed by the co-crystallized residues are separated by an immunodominant region. This region contains a short loop held together by an intramolecular disulphide bond, a structure that is common to all lentiviral transmembrane proteins⁵.

The crystal structures^{1,2} show that three molecules of each peptide fragment are present in the stable configuration, supporting the view that the glycoprotein complex on the HIV-1 envelope is trimeric⁶. After years of uncertainty caused by biochemical evidence indicating that gp41 was a tetramer⁷, the structure of the simian immunodeficiency virus (SIV) matrix protein, p17, showed that the cavity into which the internal moieties of gp41 molecules presumably fit is, in fact, triangular⁸.

The crystallized helices represent what is thought to be either the form of gp41 that exists after HIV–cell fusion has been triggered or, perhaps, a fusion intermediate. But much can be inferred about the pre-fusion structure. An analogy is that of an old-fashioned mousetrap, with a needle (the fusion peptide) attached to the end that is under tension. The helical peptide fragments represent the two halves of the trap, separated by the disulphide loop at the hinge. Triggering the mousetrap brings both helices together (in an antiparallel configuration), and fires the fusion peptide ‘needle’ down into the target cell membrane. So the profound conformational change in gp41 that occurs during fusion releases stored energy, which presumably helps to initiate the energetically unfavourable coalescence of the viral and cell membranes (Fig. 1).

What triggers the mousetrap’s spring? For influenza viruses, a change in pH after the virion has entered the cell (by endocytosis) activates fusion of the viral and endosomal membranes³. But HIV-1 fuses at the plasma membrane without a pH change, in a process that involves binding of the virus to CD4, a molecule that is found on the target cell surface, and to the co-receptors that are necessary for efficient fusion (CCR5, CXCR4, and others)⁹. The binding of HIV-1 to soluble CD4 (in the absence of co-receptors) is enough to cause massive conformational changes in the viral envelope¹⁰. More-

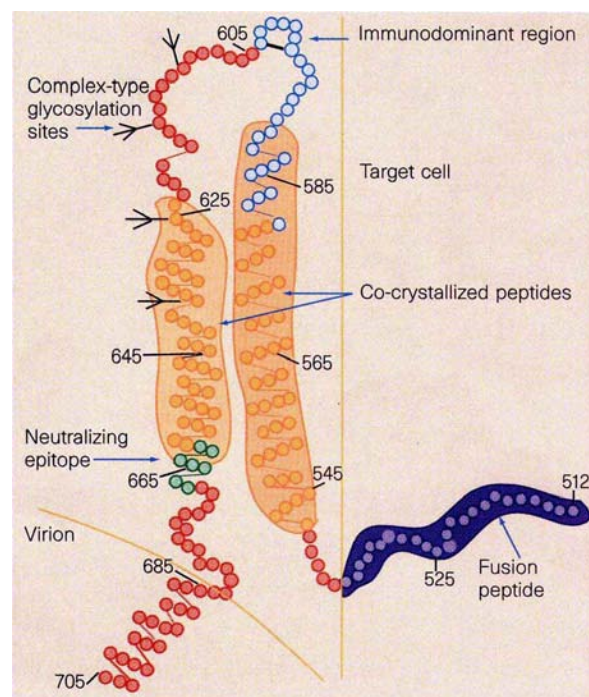


Figure 2 Secondary structure of the post-fusion conformation of gp41. The two α -helical regions that have been co-crystallized by Weissenhorn *et al.*¹ are depicted. Sequences of note are the fusion peptide (residues 512–535; mauve), the immunodominant region (residues 579–604; blue), the disulphide loop (residues 598–604) and the 2F5 neutralization epitope ELDKWAS (residues 662–668; green).

over, soluble CD4 can activate fusion of SIV (ref. 11), HIV-2 (ref. 12) and some primary forms of HIV-1 (ref. 13) — hence the concept of CD4-activated, virus–cell fusion¹⁴.

But how, precisely, are co-receptors involved, and why are they needed? A clue comes from observations that a strain of HIV-2 can fuse efficiently in a CD4-independent, co-receptor-dependent manner¹⁵. Another strain of HIV-2 can do the same thing in the presence of soluble CD4 (ref. 13). Perhaps these viruses have evolved a structure in which fusion can be partially activated in the absence of CD4 (ref. 15) or in the presence of soluble CD4 (ref. 13), although it is fully triggered only through an interaction with a co-receptor. To extend the analogy, HIV binding to CD4 is baiting the mousetrap with cheese — and the co-receptors are the mice. But some mice either don't need the cheese to spring the trap¹⁵, or else they bring it with them¹³.

In molecular terms, the primary interaction of most strains of HIV-1, HIV-2 and SIV with the cell is through CD4. The interaction induces structural changes in the virus, to create (or expose) the co-receptor-binding site(s) on gp120 (Fig. 1)^{16,17}, which is another envelope glycoprotein that is formed from the gp160 precursor. In the absence of a co-receptor, the altered configuration of the viral envelope is unstable, and fusion can proceed prematurely, inactivating the virus (rapidly for HIV-1 laboratory strains^{10,14} and slowly, perhaps, for some SIV and HIV-2 strains^{11,13–15}). But if a co-receptor successfully interacts with gp120, the structural changes in gp41 are rapidly triggered, and fusion proceeds. The co-receptors are considered to be relatively flush with the cell membrane⁹, so for the fusion peptide to reach the cell membrane (rather than firing

harmlessly into space), attachment of HIV-1 to the co-receptors might be necessary. Both geometry and time are likely to be important in determining whether successful or abortive fusion occurs, once the process is initiated by the binding of (soluble) CD4.

Other puzzles regarding the structure and function of gp41 remain. For example, why does substitution of a single residue in the immunodominant region (Fig. 2) create a mutant virus that resists anti-gp120 antibodies¹⁸? And how do gp41 peptides (DP178 and DP107) with antiviral activity act as trans-dominant inhibitors of fusion¹⁹? Perhaps antiviral drugs could be designed to fit in the cavity in gp41, just as Weissenhorn *et al.*¹ and Chan *et al.*² have proposed that the carboxy-terminal peptide DP178 does. It would also be interesting to know whether there is cooperativity among the gp41 trimers during fusion, and how this might occur. Is fusion prevented by a human monoclonal antibody (2F5), which binds an epitope near the end of one of the crystallized fragments²⁰ and, if so, how? The 2F5 monoclonal antibody is currently the only broadly neutralizing human anti-gp41 monoclonal antibody known — is this because in the native gp120–gp41 complex, most of gp41 is inaccessible to antibody binding²¹, or because most antibodies are raised to the post-fusion form of gp41, so they never recognize the more relevant native structure?

Vaccine researchers have now realized that the native structure of the HIV-1 envelope (as it exists on the virion) needs to be better understood, both for designing effective immunogens and for analysing the consequences of immunization. Structural studies, such as those by Weissenhorn *et al.* and Chan *et al.*, may yield dividends in the

future, by furthering our understanding of the process that we seek to stop: HIV-cell fusion. □

James Binley and John P. Moore are at the Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York 10021, USA.

- Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. *Nature* **387**, 426–430 (1997).
- Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. *Cell* **89**, 263–273 (1997).
- Carr, C. C. & Kim, P. S. *Cell* **73**, 823–832 (1993).
- Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. *Nature* **371**, 37–43 (1994).
- Gallagher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C. & Montelaro, R. C. *AIDS Res. Hum. Retroviruses* **5**, 431–440 (1989).
- Lu, M., Blacklow, S. C. & Kim, P. S. *Nature Struct. Biol.* **2**, 1075–1082 (1995).
- Pinter, A. et al. *J. Virol.* **63**, 2674–2679 (1989).
- Rao, Z. et al. *Nature* **378**, 743–747 (1995).
- D'Souza, M. P. & Harden, V. A. *Nature Med.* **2**, 1293–1300 (1996).
- Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. *Science* **250**, 1139–1142 (1990).
- Allan, J. S., Strauss, J. & Buck, D. W. *Science* **247**, 1084–1088 (1990).
- Sullivan, N., Sun, Y., Li, J., Hofmann, W. & Sodroski, J. *J. Virol.* **69**, 4413–4422 (1995).
- Reeves, J. D. et al. *Virology* **231**, 130–134 (1997).
- Allan, J. S. et al. *Science* **252**, 1322–1323 (1991).
- Endres, M. J. et al. *Cell* **87**, 745–756 (1996).
- Wu, L. et al. *Nature* **384**, 179–183 (1996).
- Trkola, A. et al. *Nature* **384**, 184–186 (1996).
- Wilson, C. et al. *J. Virol.* **64**, 3240–3248 (1990).
- Wild, C., Oas, T., McDaniel, C., Bolognesi, D. & Matthews, T. *Proc. Natl Acad. Sci. USA* **89**, 10537–10541 (1992).
- Muster, T. et al. *J. Virol.* **67**, 6642–6647 (1993).
- Sattentau, Q. J., Zolla-Pazner, S. & Poignard, P. *Virology* **206**, 713–717 (1995).

Intergalactic medium

Peering between the clouds

Peter Jakobsen

It is not very often that the Hubble Space Telescope is pointed at a single, unphotogenic object for more than 18 orbits in total. In the May issue of the *Astronomical Journal*, Craig Hogan, Scott Anderson and Martin Rugers' report on a lengthy spectroscopic visit to the high-redshift ($z = 3.28$) quasar Q0302–003. This is the most remote of only four quasars discovered so far that are bright enough at ultraviolet wavelengths to reveal absorption from singly ionized helium in intergalactic space along the line of sight. The closer view of the helium absorption towards Q0302–003 provided by these new observations has provided further evidence that intergalactic space at high redshift (and thus in the early Universe) is not entirely empty, but filled with a tenuous plasma that is kept highly ionized by radiation from quasars — as in some simulated universes (Fig. 1).

On its journey to Earth, the light from every quasar acquires the spectral imprint of intervening intergalactic neutral hydrogen (Fig. 2). The so-called Lyman forest of narrow Lyman-alpha absorption lines from hydrogen has been studied intensively for some 25 years using the largest ground-based telescopes, and more recently the Hubble Telescope and other space observatories. Intergalactic space appears to be peppered with tenuous clouds of possibly primordial gas that have not yet condensed into galaxies.

The intergalactic gas is bathed in a sea of ionizing radiation from quasars, which are much more numerous at high redshift. Because the quasar flux keeps at least 99.9 per cent of the intergalactic gas ionized, the observed neutral hydrogen, which causes the absorption, is only the tip of the iceberg in terms of the total baryonic mass present. An open question, however, is whether more tenuous plasma exists between the clouds. Such a component would reveal its presence as continuous redshift-smeared absorption

between the forest lines, an effect that is not evident.

Singly ionized helium (He^+) is much more resilient to further ionization from the bombarding quasar radiation than is hydrogen, requiring four times the ionization energy. Because of this, low-density ionized intergalactic gas is more easily detected in absorption by He^+ than by neutral hydrogen. But because the main transition of He^+ has such a short wavelength (30.4 nm), it can only be observed by space-borne ultraviolet telescopes at very high redshift, and thus in far less detail than hydrogen absorption, which can be seen using far larger telescopes from the ground. So the interpretation of the helium data must be guided by matching ground-based observations of the hydrogen Lyman-alpha forest.

Since the first detections^{2,3}, it has been clear that the strength of the He^+ absorption requires an additional source of He^+ beyond the conventional Lyman-forest clouds, in many lower-density clouds and/or in a truly diffuse intergalactic medium. The subsequent discovery of very weak Lyman-forest systems⁴ provided a convincing source for at least part of the extra He^+ , leaving much less room for a diffuse component.

The new high-resolution ultraviolet spectrum of Q0302–003, obtained by Hogan and co-workers¹ using the Goddard High Resolution Spectrograph, has now raised the ante further, by revealing several subtle details that suggest that there may yet be diffuse gas hiding in the helium data. As it happens, Q0302–003 is one of only a few quasars known to display a clearing, or 'void', in the Lyman forest along its line of sight⁵. The new spectrum shows that this void is not mimicked in the He^+ absorption, the implication being that the He^+ opacity is dominated by gas that is more diffuse than the Lyman-forest systems. Moreover, at redshifts very close to the quasar (corresponding to distances of less than about seven million parsecs), this diffuse component appears to have been ionized into invisibility by the extra radiation coming from Q0302–003, leaving only the denser forest systems, several of which are convincingly detected from their He^+ lines.

The new data do not tell whether the diffuse helium absorption is due to truly diffuse intergalactic gas or to unresolved clouds even weaker than those now detectable. However, if one is to believe recent simulations of the behaviour of the gaseous component of the Universe, this may be something of a moot point. Several groups^{6–8} have reproduced theoretically much of the

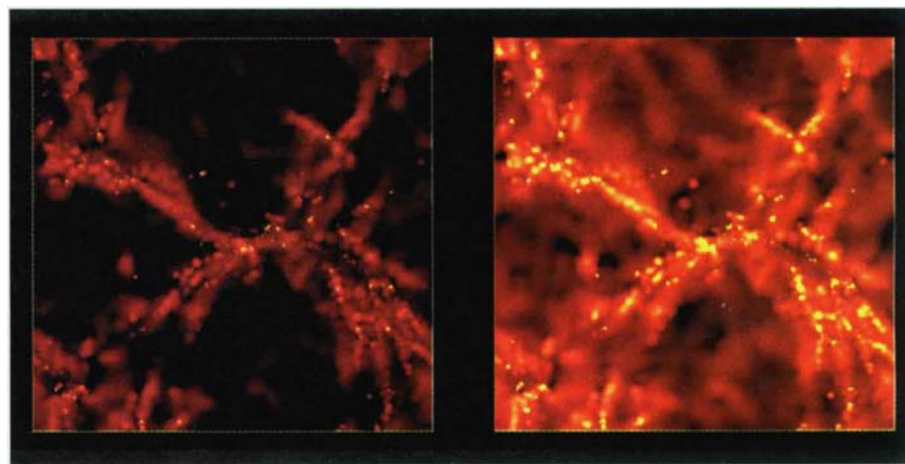


Figure 1 The thinnest clouds. The intergalactic medium in a simulated universe^{8,11} dominated by cold (that is, non-relativistic) dark matter. The distribution of neutral hydrogen (left) and singly ionized helium (right) is shown in a thin slice through the centre of the simulation cube, which is 22 megaparsecs on a side. Quasar light shining through the box is absorbed by the gas to different degrees, with intermediate absorption shown in red and the highest in yellow. Black corresponds to absorption of 5 per cent or less. Helium absorption is prominent and measurable in low-density regions that are nearly transparent to neutral hydrogen. Helium absorption can be used to probe diffuse structures that are currently inaccessible by any other means. (From ref. 11.)

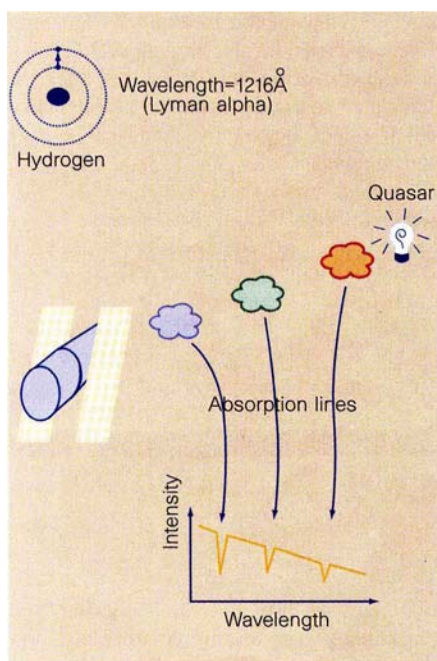


Figure 2 Clouds of gas between us and a distant quasar all absorb light at the same (Lyman-alpha) wavelength in their own rest-frame, but because they have different redshifts, many separate absorption lines are seen — the Lyman forest.

diverse structure of the Lyman forest, using the popular cold-dark-matter model of galaxy formation. This work has led to a blurring of the distinction between the forest clouds and the intergalactic medium: in the cold-dark-matter model, the hard work of forming structure in the Universe is done by the dark matter, which carries essentially all of the mass, and the normal baryonic component is merely 'along for the ride', slowly collecting in the potential wells created by the gravitational collapse of the dark matter. In this picture, there are no well-defined forest 'clouds' as such, but rather a single, continuous medium filling the volume of intergalactic space, yet showing large, but smooth, variations in density and velocity along any given line of sight (Fig. 1). The seemingly continuous nature of the He^+ absorption seen in the new spectrum of Q0302-003 is consistent with this.

The ionization levels inferred from the Q0302-003 spectrum are also consistent with the notion that most normal matter at high redshift hides in the ionized component of the Lyman forest. Recent measurements of the deuterium content of the Lyman-forest clouds^{9,10}, applied to theories of nucleosynthesis in the Big Bang, constrain the baryonic content of the Universe to no more than a few per cent of the critical density (required for a just-closed Universe) — an amount that can readily be found in the helium data.

We can hope to see further improvements in the detection of intergalactic He^+ absorption in the coming years, once STIS — the advanced ultraviolet spectrograph that was

successfully substituted for the Goddard Spectrograph during the recent Hubble Telescope servicing mission — comes on line. □

Peter Jakobsen is in the Astrophysics Division, Space Science Department of ESA, ESTEC, 2200 AG Noordwijk, The Netherlands.

1. Hogan, C. J., Anderson, S. F. & Rugers, M. H. *Astron. J.* **113**, 1495–1505 (1997).
2. Jakobsen, P. *et al. Nature* **370**, 35–39 (1994).
3. Davidsen, A. F., Kriss, G. A. & Zheng, W. *Nature* **380**, 47–49 (1996).

4. Songaila, A., Hu, E. M. & Cowie, L. L. *Nature* **375**, 124–126 (1995).
5. Dobrzycki, A. & Bechtold, J. *Astrophys. J.* **377**, L69–L72 (1991).
6. Cen, R., Miralda-Escudé, J., Ostriker, J. P. & Rauch, M. *Astrophys. J.* **437**, L9–L12 (1994).
7. Zhang, Y., Anninos, P. & Norman, M. *Astrophys. J.* **353**, L57–L60 (1995).
8. Hernquist, L., Katz, N., Weinberg, D. H. & Miralda-Escudé, J. *Astrophys. J.* **457**, L51–L55 (1996).
9. Songaila, A., Cowie, L. L. & Wampler, E. J. *Nature* **385**, 137–139 (1997).
10. Tytler, D., Fan, X. M. & Burles, S. *Nature* **381**, 207–209 (1996).
11. Croft, R. A. C., Weinberg, D. H., Katz, N. & Hernquist, L. *Astrophys. J.* (submitted).

Avian origins

A new missing link

Lawrence M. Witmer

Among evolutionary debates, perhaps only the origin of humans has attracted as much attention as avian origins. Less than two years after Darwin's *Origin of Species* hit the shelves, a spectacular 'missing link' between birds and reptiles was unearthed in Germany: the famous 'feathered reptile' *Archaeopteryx lithographica*, which immediately became central not only to the debate on the evolution of birds, but to the debate on evolution itself. Now, on page 390 of this issue, Novas and Puerta¹ announce the discovery of a new intermediate form, which may help to fill in the details of the functional transition between reptiles and birds.

Since the 1860s, scientists have debated which of the various reptilian groups is genealogically the closest to *Archaeopteryx*. Most of the main groups of archosaurs have been linked with birds at one time or another but, for the past 20 years², there has been a broad consensus (although not without dissent³) that birds are related to a group of small, bipedal, carnivorous dinosaurs known as theropods. Among theropods, dromaeosaurid coelurosaurs such as

Deinonychus and *Velociraptor* are usually regarded as the closest relatives of *Archaeopteryx* and other birds^{4,5}. But a troubling (if narrowed) morphological gap has remained, impeding attempts to tease apart the details of the functional transition.

Novas and Puerta¹ now describe a non-avian theropod dinosaur which they call *Unenlagia comahuensis*, from the early Late Cretaceous (90 million years ago) in Patagonia, Argentina. *Unenlagia* slots in cladistically between dromaeosaurids and *Archaeopteryx*. It was a relatively small theropod, perhaps only two-thirds the size of *Deinonychus*, and, typical of these small, fragile theropod skeletons, *Unenlagia* is known from only a handful of bones. Nevertheless, Novas and Puerta have selected an apt name in that *Unenlagia* means 'half bird', and it indeed combines an almost even mixture of primitive coelurosaurian and derived avian characters. For example, the *Unenlagia* ischium (Fig. 1) has the typical triangular obturator process that is found in most advanced coelurosaurs, as well as a process that is found only in *Archaeopteryx* and other

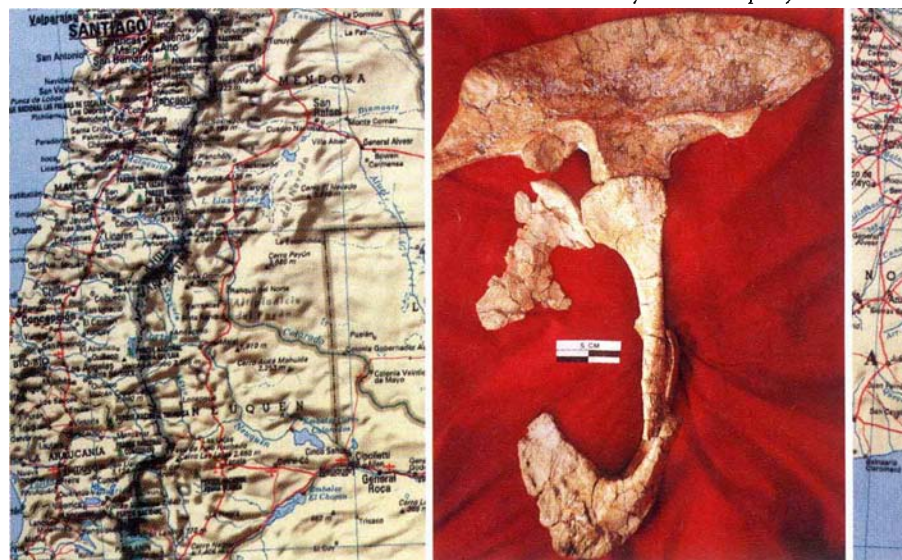


Figure 1 Novas and Puerta¹ have used a small number of bones to build up a picture of a new intermediate form (*Unenlagia comahuensis*) between reptiles and birds. The bones were found in the Patagonian badlands (Neuquén Province), and the *Unenlagia* pelvis is shown. The pelvis has features that are found both in birds and in most advanced coelurosaurs. (Photo courtesy of F. E. Novas.)

birds. The ilium has the usual excavation for a certain muscle found in coelurosaurs, yet it also has an extensive inner wall to the hip socket, much as in *Archaeopteryx* and other basal birds. This was unexpected because non-avian dinosaurs almost always have a fully open socket. The pubis is at a dromaeosaurid grade of retroversion, meaning that it is directed more towards the rear than in more primitive coelurosaurs, but not as far as in *Archaeopteryx* and other birds. Moreover, the pubis has a large coelurosaurs-like distal expansion, or 'boot', yet, as in all birds, *Unenlagia* has totally lost the primitive forward 'toe' of the boot. So *Unenlagia* is a true mosaic, begging the question of where to draw the line between what is, or is not, a 'bird'.

The one feature that will raise the most controversy is the orientation of the shoulder joint in *Unenlagia*. Novas and Puerta believe that the joint surface faces laterally, as in *Archaeopteryx* and other birds. But this would be totally unlike non-avian dinosaurs, in which the shoulder socket faces more down and back. With a laterally facing joint surface, *Unenlagia* could raise its forelimb extensively, producing a full upstroke in preparation for a long, down-and-forward 'flight stroke'. Furthermore, the orientation is another step towards the wing-folding mechanism which, in birds, allows the long flight feathers to be safely tucked away. So according to Novas and Puerta, *Unenlagia* shows that many of the features that we associate with avian flight evolved in a terrestrial context — and perhaps flight evolved 'from the ground up', without an intervening arboreal, gliding phase.

This interpretation will be controversial because the orientation of the shoulder socket is dependent on the orientation of the scapula, which is not fixed to the rib cage, but rather 'floats' in muscle. If the scapula is placed in an avian position, the socket faces laterally, whereas in a more dromaeosaurid-like position, the orientation of the socket is less avian. But some features of the scapula of *Unenlagia* support an avian orientation. For example, the shaft of the scapula is twisted to lie on the back, as in most birds. Perhaps more importantly, there is a prominent, projecting acromion process that is very similar to that found in *Archaeopteryx* and other birds, but which is not found in non-avian theropods. These features may signal the restructuring of the shoulder girdle along avian lines.

Might we then suppose that *Unenlagia* had feathers? Certainly no feathers were preserved, but given the nature of the rocks in which *Unenlagia* was found, this is not surprising. Although it is tempting to speculate that feathers, and perhaps even incipient flight, honed these bird-like attributes of *Unenlagia*, there is as yet no evidence for this. Novas and Puerta¹ are appropriately cautious, regarding *Unenlagia* as a relic descendant of a lineage that diverged just

before the fully feathered and flying line that leads to *Archaeopteryx* and true birds. In fact, some of the common features may ultimately have been co-opted for flight, but they could have originated in a completely different functional context. What could this context have been? Unfortunately, forelimb function in advanced coelurosaurs is poorly understood, particularly with regard to joint mobility and excursion. Until such biomechanical questions are resolved, an arboreal climbing model may be as valid as a terrestrial predation model in explaining the transition to birds. But regardless of the final

interpretation of the functional transition, *Unenlagia* is a critical testament of the phylogenetic transition, and it is only after the phylogenetic players are better known that the drama of the adaptive story will be truly comprehensible. □

Lawrence M. Witmer is in the Department of Biological Sciences and College of Osteopathic Medicine, Ohio University, Athens, Ohio 45701, USA.

1. Novas, F. E. & Puerta, P. E. *Nature* **387**, 390–392 (1997).
2. Ostrom, J. H. *Biol. J. Linn. Soc.* **8**, 91–182 (1976).
3. Feduccia, A. *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven, CT, 1996).
4. Holtz, T. H. *J. Paleontol.* **68**, 1100–1117 (1994).
5. Chiappe, L. M. *Nature* **378**, 349–355 (1995).

Host-guest chemistry

Crowns get organized

Thomas E. Mallouk

By combining organic and inorganic building blocks, composite materials can be created that embody the most useful features of both. In a recent article¹, Baolong Zhang and Abraham Clearfield describe a lamellar solid made of zirconium phosphonate sheets, which are held apart by crown ether molecules. The crystalline inorganic framework controls the juxtaposition of these organic hosts, and the access of other molecules and ions, making the whole a promising structure for designed sensors, separations media and catalysts.

Small guest molecules bind to macromolecular hosts in many important chemical processes. This binding is a prerequisite for the function of enzyme active sites, chromatographic media and many heterogeneous catalysts. As with Cinderella's glass slipper, a good fit is needed for the guest to bind tightly to the host: oversized guests do not fit, and small ones come loose on the palace steps. Some inorganic solids, such as zeolites and clays, contain cavities that are just right, and these materials also tend to be stable in chemically aggressive environments. They are of great interest in separations technologies, for example^{2,3} in the extraction of radioactive Cs⁺ and Sr²⁺ from liquid wastes containing a vast excess of Na⁺. Unfortunately, crystalline solids are not always amenable to design — the strong chemical bonds that hold together inorganic

materials are also structure-directing, and make dense structures much more common than open ones.

In contrast, organic hosts (like slippers) can be made in many sizes, and their names (crowns, cavitands, cryptands, torands, cyclams, and so on) describe a rich variety of shapes. So one can usually be chosen or designed to select the appropriate guest from a mixture. Crown ethers are so called because they resemble a crown, the points being oxygen atoms that attach themselves to metal ions and neutral solutes. They have been widely used in chemical analysis⁴ and separations⁵, and their high affinity for cations has more recently stimulated applications in phase-transfer catalysis. Although cation-selective hosts are most common, molecules that bind anions have also been developed^{6,7}.

But the common problem with organic hosts is that, to be recovered easily after a separation or catalysis, they must be anchored to insoluble supports such as polymers^{8,9} or silica¹⁰. Attached randomly to the surface of these materials, the host molecules experience a range of microenvironments, and there is no control over their juxtaposition, which might otherwise be used to advantage in applications that require cooperative binding. Such materials are also difficult to characterize precisely, because of their heterogeneity and amor-

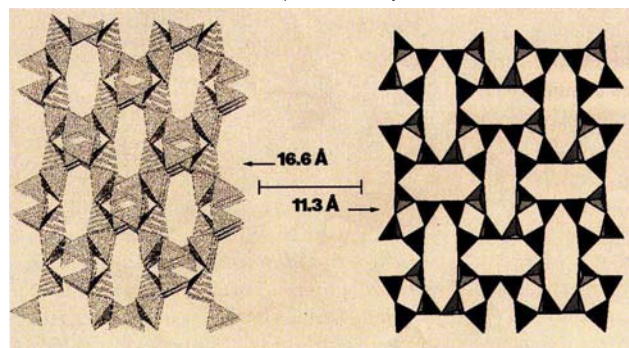


Figure 1 Molecular minerals. Like oxygen, the pyrimidine ligand makes two angled bonds, so the copper-pyrimidine network (left) resembles that of a feldspar, MAISI₈O₈ (right). By mimicking minerals, active organic molecules can be assembled into stable, predictable structures.

phous nature. These new lamellar compounds, on the other hand, present to reactive guests an ordered array of host molecules (the crown ethers), with well-defined access pathways. The combination of a highly selective organic host with stable inorganic scaffolding makes these materials attractive for use in harsh environments, such as the separation of radioactive cations from aqueous nuclear waste.

Modular synthesis, as used by Zhang and Clearfield¹, is one way to combine a tailored organic host, or any small molecule with a particularly useful function, with well-defined inorganic scaffolding. The idea is simple: mix metal ions with a preference for a particular coordination geometry (such as octahedral or tetrahedral) with connecting organic ligands. Under the right conditions — and it is important to control both the kinetics and thermodynamics of the assembly process — a crystalline network will nucleate and grow. The topology of the network can often be predicted. For example, most transition metals combine with phosphonates to produce layered inorganic-organic networks, because the phosphonate group is shaped like a tripod¹¹. Coordination solids that are topologically related to mineral structures are common: tetrahedral metals and linkers, for example, produce a diamond network¹², whereas octahedra and triangles give a structure¹³ resembling rutile (TiO₂).

Coordination solids can be made under very mild conditions, even at room temperature. Topologically similar, dense inorganic solids are usually prepared under more extreme conditions. For example, feldspars (Fig. 1) are igneous aluminosilicates, which crystallize at high pressure and temperature. And the building units are large — the pyrimidine (C₄H₄N₂) ligand, for example¹⁴, takes up more room than an oxygen atom — so coordination networks are usually 'expanded' (Fig. 1). In favourable cases the expansion creates an open framework. But if the void volume becomes too great, we are reminded of nature's aversion to a vacuum: two or more interpenetrating networks form, and a dense solid obtains¹⁵.

The design of coordination solids is a young field. The rules of lattice topology and interpenetration are still being learned, and the synthesis of hydrolytically stable networks is a persistent challenge. Most experiments to date have focused on preparing stable solids with molecule-size cavities and access pathways¹⁵. These features are desirable in practical applications that rely on host-guest interactions, but by themselves are not enough. A new development in this area is the engineering of 'hinged' networks, which adapt themselves to the inclusion of different-size guests¹⁶. Add to these networks the affinity and selectivity of organic hosts, and then the possibilities for designed sen-

sors, molecular sieves, separations media, and catalysts become interesting. □

Thomas E. Mallouk is in the Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

1. Zhang, B. & Clearfield, A. *J. Am. Chem. Soc.* **119**, 2751–2752 (1997).
2. Komarneni, S. & Roy, R. *Science* **239**, 1286–1288 (1988).
3. Paulus, W. J., Komarneni, S. & Roy, R. *Nature* **357**, 571–573 (1992).
4. Kolthoff, I. M. *Anal. Chem.* **51**, 1R–22R (1979).
5. Izatt, R. M. et al. *J. Membrane Sci.* **9**, 263–271 (1981).
6. Dietrich, B. *Pure Appl. Chem.* **65**, 1457–1464 (1993).
7. Iverson, B. L. et al. *J. Am. Chem. Soc.* **118**, 1608–1616 (1996).
8. Regen, S. L. *Angew. Chem. Int. Ed. Engl.* **18**, 421–429 (1979).
9. Hodge, P. & Sherrington, D. C. *Polymer-Supported Reactions in Organic Synthesis* (Wiley, New York, 1980).
10. Kimura, K., Nakajima, N. & Shono, T. *Anal. Lett.* **13**, 741–750 (1980).
11. Cao, G., Hong, H.-G. & Mallouk, T. E. *Acc. Chem. Res.* **25**, 420–427 (1992).
12. Hoskins, B. F. & Robson, R. *J. Am. Chem. Soc.* **111**, 5962–5964 (1989).
13. Batten, S. R., Hoskins, B. F. & Robson, R. *J. Chem. Soc. Chem. Comm.* 445–447 (1991).
14. Keller, S. W. *Angew. Chem. Int. Ed. Engl.* **36**, 247–248 (1997).
15. Yaghi, O. M., Li, G. & Li, H. *Nature* **378**, 703–706 (1995).
16. Russell, V. A., Evans, C. C., Li, W. & Ward, M. D. *Science* **276**, 575–579 (1997).

Evolutionary ecology

Making life simpler

H. C. J. Godfray

There is enormous variation among species of plants and animals in lifespan, age at first reproduction, the number of offspring produced at each breeding attempt, offspring size and timing of senescence. Making sense of this diversity, and explaining how it has evolved, is the field of evolutionary ecology called life-history theory. Over the past ten years, Charnov¹ has argued that it is possible to simplify the often complex arguments of life-history theory and abstract desirable general results, but without losing the power to explain actual patterns in nature. On page 393 of this issue² he now shows how the basic equation for the darwinian fitness of an organism with overlapping generations can be reduced to a series of very simple invariant relationships.

If we know the environment in which a plant or animal lives, and if we understand the constraints and trade-offs that influence how an organism allocates resources to competing activities, it is possible to predict the life history that maximizes fitness. Our understanding of life-history evolution can then be tested by comparing the predictions with the observed biologies of real plants and animals. The modern field of life-history theory was defined 20 years ago by Stearns³, who in an influential review described the trade-offs and constraints that can influence the many components of an organism's life-history strategy, and by Charlesworth⁴, who showed how to calculate darwinian fitness in populations with overlapping generations that may be changing in numbers over time. Because of the hydra-headed nature of the subject, and because the calculation of fitness in populations with overlapping generations can be complex (fitness is defined implicitly by the solution of a non-algebraic expression), life-history theory has a reputation for being a difficult subject with few general results.

Charnov's attack on the complexities of life-history theory has several steps. First he simplifies the calculation of fitness by assuming that the population is stationary,

neither growing nor diminishing in size. When a population is growing in size, offspring produced early in life are of higher value than those produced later in life because the former reach reproductive age and have offspring before the latter. Early reproduction is favoured for exactly the same reason that it makes sense not to delay putting money into a bank account that pays interest. But in a stationary population the non-algebraic complexities mentioned above disappear and darwinian fitness is simply the number of offspring produced over the organism's lifespan, a quantity traditionally referred to as R_0 .

Can he get away with this? The answer depends on the organism. All persistent populations are regulated by some form of density dependence. If the density dependence acts in a simple manner (for example, by killing excess offspring before they reach reproductive maturity), and if population densities fluctuate little from generation to generation, then R_0 will accurately map onto fitness. But if populations fluctuate violently in size — think of aphids or lemmings — then R_0 will only poorly approximate fitness.

A plant or animal's life history is normally described by two series of numbers known in the trade as l_x and m_x , respectively the probability of living until age x and the number of offspring produced at age x . The units in which x is measured depend on the organism — years for trees or large mammals, days for many insects, hours for some microorganisms. In a stationary population, fitness (R_0) is simply the sum of $l_x m_x$ taken over all ages.

Charnov's second strategy to simplify life-history theory is to ignore most of the details of how l_x and m_x change with age and express fitness as the product of three numbers: the probability of reaching the age of reproductive maturity, the average number of offspring produced per unit time when mature, and the average reproductive life of the organism. Going further, one can combine two of these three numbers to express fitness in terms of the product of just two

numbers. This can be done in different ways, and Charnov discusses several alternatives. For example, multiplying the average number of offspring produced per unit time by expected reproductive lifetime allows fitness to be expressed as the product of the probability of surviving until the age of reproductive maturity and expected lifetime reproductive success.

What is gained by this paroxysm of reductionism? Suppose fitness is the product of two quantities *A* and *B*. The organism can increase its fitness by allocating resources (energy, time, and so on) to either quantity. But resources are limiting and there will be a trade-off: allocating more resources to obtain higher fitness through *A* will be accompanied by a reduction in fitness through *B*. Where on this trade-off curve will the optimal life history lie? The answer is given by elementary calculus and is simply where the slope of the trade-off curve measured on a logarithmic scale is -1 ; in symbols, where $\partial \ln(A)/\partial \ln(B) = -1$.

A number of authors have analysed specific problems in life-history theory and have come up with results that are related to Charnov's. What is new is the argument that the same basic pattern underlies so many of the classical questions in life-history theory. The simplicity has, of course, been gained at a price: many of the exquisite details that characterize biological diversity have been averaged away, and the trade-offs between

the portmanteau quantities that remain may be very difficult to estimate from data. But it is notable that sex-allocation theory⁵ (which explains how animals allocate resources to male and female functions), perhaps the single most successful application of optimization models in evolutionary biology, is based on the analysis of a simple product expression for fitness (the product of the fitness benefits of producing sons and daughters).

Life-history theory is currently being enlivened by a number of new ideas. For example, as described in a recent News and Views article⁶, Kozłowski and Weiner⁷ have fused optimality arguments and body-weight scaling of life-history variables to predict the broad patterns of mammalian life histories, a subject also tackled by Charnov¹. Which of these new approaches ultimately survive will depend on their success in explaining empirical patterns in the data: an appropriately darwinian criterion. □

H. C. J. Godfray is in the Department of Biology and the NERC Centre for Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.

1. Charnov, E. L. *Life History Invariants* (Oxford Univ. Press, 1993).
2. Charnov, E. L. *Nature* **387**, 393–394 (1997).
3. Stearns, S. C. *Q. Rev. Biol.* **51**, 3–46 (1976).
4. Charlesworth, B. *Evolution in Age-Structured Populations* 2nd edn (Cambridge Univ. Press, 1994).
5. Charnov, E. L. *Sex Allocation Theory* (Princeton Univ. Press, 1982).
6. Purvis, A. & Harvey, P. H. *Nature* **386**, 332–333 (1997).
7. Kozłowski, J. & Weiner, J. *Am. Nat.* **149**, 352–380 (1997).

Plant biology

Fixing a symbiotic circle

Allan Downie

Many bacteria contain 'mega'-plasmids (extrachromosomal DNA circles), some of which are larger than the entire genome of *Mycobacterium genitalium* (580 kilobase-pairs; kb). To find out what kinds of genes are present on these plasmids, Freiberg *et al.*¹ have sequenced a 536-kb plasmid from a strain of *Rhizobium*. On page 394 of this issue they report that this plasmid, which is not required for free-living bacterial growth, has a very different range of genes from those that have been identified as a result of other bacterial genome sequencing projects.

Bacterial plasmids provide a reservoir of 'adaptive' genes which, although dispensable, can enable a bacterium to adopt a specialized lifestyle. The *Rhizobium*–legume symbiosis is a classic example — rhizobia grow slowly for long periods in soil, but if they infect a compatible legume they can grow rapidly; successful infection (Fig. 1) by a single bacterium can lead to the formation of a nitrogen-fixing nodule on the root of the legume, containing over 10^8 bacterial progeny. So the potential

for genetic selection is great.

The genes that are required to form these symbiotic nodules are often found on plasmids. Many bacteria contain different plasmids that determine characteristics of medical, economic and environmental importance: for example, virulence genes of animal and plant pathogens; the ability to degrade environmentally toxic compounds (including oils and phenolics); the synthesis of extracellular enzymes (used in various industrial and food-related processes); the production of antibiotics and, of course, antibiotic-resistance genes. Many of these characteristics have been studied in great detail, but Freiberg *et al.* are the first to sequence one of the very large plasmids.

Why sequence the plasmid from this particular *Rhizobium* strain? It was chosen because it has an interesting history. The strain was first isolated in New Guinea, and it was found to induce nodules on the non-leguminous tree *Parasponia andersonii*, as well as on over 100 genera of legumes (many rhizobia can nodulate only a few genera). This broad host range is correlated with its

ability to make an unusually large variety of nodulation (Nod) factors² — oligomers of four or five β 1–4-linked *N*-acetylglucosamine residues carrying an *N*-linked fatty acid and a variety of substituent groups. Legumes can recognize Nod factors at concentrations as low as 10^{-12} M, inducing deformation of root hairs and expression of plant genes. At higher concentrations, Nod factors can induce nodule morphogenesis in the absence of bacteria³. In general, the precise structures of the Nod factors determine which legumes can be nodulated³.

In most rhizobia, around four to six operons are involved in legume recognition. Each is preceded by a characteristic promoter sequence known as the *nod* box. Freiberg *et al.*¹ studied the plasmid pNGR234a, and they found 17 potential *nod* boxes, at least 11 of which are symbiotically active. The biochemical functions of several of the *nodulation* genes within these operons (particularly those involved in the biosynthesis of nodulation factors) are already known³. But hitherto undiscovered genes downstream of the *nod* boxes include genes that may be involved in polysaccharide biosynthesis and protein secretion, and a putative transmembrane transporter.

Rhizobia also contain genes that are involved in nitrogen fixation. Unlike the *nod* genes on pNGR234a, these (*nif* and *fix*) genes tend to be clustered together. Freiberg *et al.* identified genes that may be involved in electron transport, along with other nodule-expressed open reading frames of unknown function. There are also several putative *nifA*- σ^{54} -type promoter sequences, which are normally expected to be associated with

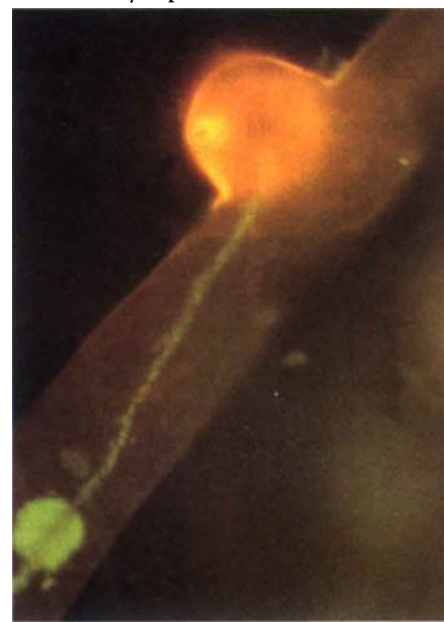


Figure 1 Root hair, showing the site of infection by the bacterium *Rhizobium*. The rhizobia are contained within the infection thread, and the nucleus that is guiding the infection process is also shown. (Figure courtesy of W. J. Broughton.)

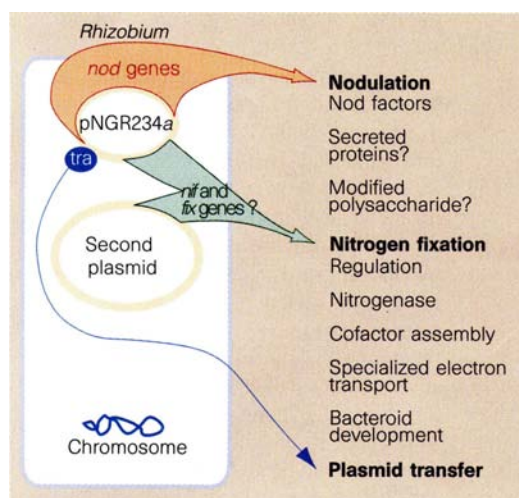


Figure 2 Freiberg *et al.*¹ have sequenced a 536-kb plasmid (pNGR234a), which is not required for free-living bacterial growth, from a strain of *Rhizobium*. Included among the genes that they have identified are the *nod*, *nif* and *fix* genes, involved in the establishment of nitrogen-fixing nodules on the roots of legumes. This *Rhizobium* strain contains a second plasmid, which may also carry genes that are necessary for an effective symbiosis.

nitrogen-fixation genes. Surprisingly, some of these sequences are upstream of genes that are probably involved in polysaccharide biosynthesis or modification. Although surface polysaccharides are unlikely to be directly involved in nitrogen fixation, there are clearly considerable changes to the bacterial cell surface during the maturation of bacteroids⁴ (bacterium-like cells found in root nodules). Within root nodules, nitrogen-fixing bacteroids are in intimate contact with the 'peribacteroid membrane' — a differentiated and modified form of the plant plasma membrane that controls the flow of metabolites between plant cytoplasm and nitrogen-fixing bacteroids. Perhaps alterations to the bacterial cell surface are necessary to maintain this crucial interface.

An advantage of having symbiosis genes on a plasmid is that, in principle, they can be transferred into previously non-symbiotic bacteria that may be better adapted to the particular soil conditions. Such transfer, which, for example, has been observed in New Zealand soils⁵, enables the plasmid genes to survive, albeit in another bacterial host. Freiberg *et al.* found a cluster of plasmid-transfer genes on pNGR234a. These are probably induced by acyl homoserine lactones in a cell-density-dependent way, analogous to that seen in *Agrobacterium* spp.^{6,7}. But surprisingly, although pNGR234a has been studied genetically for over 15 years, no one noticed that it is self-transmissible — and sequencing a 536-kb plasmid seems to be a roundabout way to find this out!

Several genes that would be expected to be involved in rhizobial symbiotic nitrogen fixation are absent from pNGR234a. Examples include the oxygen-responsive regulatory genes *fixLJ*; the regulator *fixK*; *fixNOQP*, which encode the bacteroid high-affinity cytochrome oxidase complex; the *fixABC* genes that are required for nitrogen activity; and the *fixGHIS* genes, which are thought to encode a redox-process-coupled ion pump. These genes are all conserved among diverse rhizobia⁸ and, although they

are not found in pNGR234a, they are expected to be present elsewhere. Have they moved to the chromosome or are they on the second rhizobial mega-plasmid⁹, which is thought to be half as big again? It is a pity that Freiberg *et al.* do not refer to this second plasmid, as it may also carry genes that are required for an effective symbiosis (Fig. 2).

Almost one-fifth of the total pNGR234a sequence is made up of insertion elements and mosaic sequences, and what we see can only be a snapshot of what must be a constantly evolving replicon. The insertion elements are clustered, and they tend to flank functionally important regions. So can the nitrogen-fixation genes move as a single, huge, mobile element, using flanking insertion sequences as a means of transposition? Such an event might explain how in some rhizobia (such as *Bradyrhizobium japonicum*) the symbiosis genes come to be chromosomal. The scattered distribution of the nodulation genes might reflect the relatively recent acquisition of genes that promote nodulation. The insertion elements could stimulate gene capture, either by transposition or by providing homologous regions that allow genes to be acquired from other replicons through recombination. The differences in the amounts of guanine/cytosine in the nodulation and nitrogen-fixation genes implies that they have evolved separately and, presumably, been captured from different sources.

As expected for a non-essential plasmid, pNGR234a does not contain any crucial genes that encode enzymes involved in central metabolism or DNA replication and repair. But there are 139 open reading frames (about one-third of the total sequence) with no homologies to database entries. Defining the functions of such genes could be difficult, and may have to await new database entries derived from more directed genetic or function-based approaches, possibly by others working with different bacteria. This illustrates a weakness of plasmid-genome sequencing, and highlights the point that, although genome sequencing projects reveal much information, in many ways they are a beginning — and there is still a lot of



100 YEARS AGO

Cambridge. — The Committee for promoting the admission of women to titular degrees appears to be disintegrating. On May 17 a fly-sheet was circulated by the President of Queens' (Dr. Ryle), the Registry (Mr. J. W. Clark), Mr. E. S. Roberts (tutor of Caius), and Mr. W. L. Mollison (tutor of Clare), expressing their decision to withdraw their support from the proposals before the Senate. They are now convinced the removal of the alleged grievance, felt by a comparatively small number of women, would be bought at too high a price, when considerably more than half of the resident members of the Senate are bitterly opposed to the measure, and would view it, if carried, as a grave betrayal of trust.

From *Nature* 20 May 1897.

50 YEARS AGO

The increase of milk production is one of the desiderata that is commonly accepted as necessary, and one of the equally commonly accepted ways of securing an increase is by the use of a proven sire. ... It is to this end that the Mount Hope Farm was established. The work with dairy cattle started in 1916 when the first were bought, but contagious abortion delayed the start of real work on breeding until the herd had been purged of this disease, by which time (1924) a young Guernsey herd had been acquired which was thought likely to be good. For various causes a satisfactory bull did not appear, and in 1925 a bull was bought because he ranked well according to the system of valuing dairy bulls developed at the Maine Agricultural Experiment Station. ... Misfortune followed the choice, for the bull became impotent and a new bull was selected on the Turner measure for proven sires; but this was found unsatisfactory, and this particular system is no longer used "because it has been found that it over-estimated the inheritance which the sire contributed to its off-spring"; and Mr. Parmalee Prentice and his co-workers found it conceivable that the daughters of this bull were good "not only because his influence was good but also because their dams had contributed as much as the sire" — which seems not unreasonable.

From *Nature* 24 May 1947.

scope for genetic and biochemically led research. □

Allan Downie is at the John Innes Centre, Colney Lane, Norwich NR4 7UH, UK.

1. Freiberger, C. *et al.* *Nature* **387**, 394–401 (1997).
2. Price, N. P. J. *et al.* *Mol. Microbiol.* **6**, 3575–3584 (1992).
3. Dénarié, J. *et al.* *Annu. Rev. Biochem.* **65**, 503–535 (1996).

4. Kannenberg, E. L. & Brewin, N. J. *Trends Microbiol.* **2**, 277–283 (1994).
5. Sullivan, J. T. *et al.* *Proc. Natl Acad. Sci. USA* **92**, 8985–8989 (1995).
6. Zhang, L. *et al.* *Nature* **362**, 446–448 (1993).
7. Piper, K. R. *et al.* *Nature* **362**, 448–450 (1993).
8. Fischer, H.-M. *Microbiol. Rev.* **58**, 352–386 (1994).
9. Morrison, N. A. *et al.* *J. Bacteriol.* **160**, 483–487 (1984).

Palaeontology

Progress at the K–T boundary

David Jablonski

The biotic upheaval at the end of the Mesozoic Era, often termed the end-Cretaceous or Cretaceous–Tertiary (K–T) mass extinction, was recognized by 1860¹. The past two decades, however, have seen an extraordinary wave of interdisciplinary research following the suggestion that geochemical anomalies at the K–T boundary are the signal of an extraterrestrial impact that could have triggered the extinction². Today there is widespread agreement that an asteroid or comet did indeed strike the Earth 65 million years ago, and generated the huge Chicxulub crater now buried under the Yucatan Peninsula in Mexico³. Even the handful of Chicxulub-sceptics generally accept that the impact and its associated extinction occurred⁴, but the magnitude and pattern of the biological consequences remain topics of hot debate.

A consortium of 22 palaeontologists, experts in nearly as many fossil groups, have recently produced a valuable summary of their views on the fossil record of the K–T biotic transition⁵. Some of the conclusions of that overview are controversial, but collectively the contributors clarify a number of the central issues surrounding the interval that saw the disappearance of the dinosaurs from the land, the pterosaurs and many bird groups from the sky, and the ammonites, rudist bivalves and much of the calcareous plankton from the seas.

Why has consensus been so elusive on the palaeobiological side of the K–T debates, and on mass extinctions in general? There are two main reasons. First, preservational biases limit resolution of the fine details of extinction and changing abundances in the fossil record immediately before and after an extinction event. Second, simple, unitary cause-and-effect relations are scarce in complex biological systems, so that, for example, a single 'forcing factor' might have radically different effects depending on the state of the system at the time of perturbation, and a single extinction pattern might be equally consistent with a host of potential forcing factors⁶.

To illustrate this last point, consider one simple outcome, the survival of a lineage of gastropods at the K–T boundary. This happy event might have occurred because, to name

just three possibilities among many, these snails lived in an environment that was not seriously stressed by the forcing mechanism(s) behind the extinction; or they possessed a physiology or life habit that allowed them to weather the event in a stressed environment; or they were so widespread that their overall geographical range happened to include one or more protected places that provided a refuge. Statistical approaches and large databases that can tease apart these alternatives are becoming increasingly available (and in one instance, at least, suggest that geographical range was more important in lineage survivorship than were environ-

mental preferences, physiologies or life habits⁶).

Several of the major conclusions drawn by MacLeod and colleagues⁵ from their overview of the K–T fossil record represent a discipline-wide consensus. Among them are that a short-term global extinction indeed occurred close to the K–T boundary, being especially dramatic and clear-cut in certain marine microfossil groups; that extinction intensities varied greatly among groups, from 100 per cent of ammonite species to 10 per cent of bony fish families; that the recovery of biotic diversity after the K–T event took a long time, even on geological timescales, despite the accelerated rates of diversification that might be expected as vacant niches were re-occupied in the aftermath of a significant extinction event; and that an important area for refining K–T databases is the reconstruction of evolutionary relationships, at the level of species and genera, across the extinction boundary, so that extinction intensities and selectivities can be more accurately quantified.

Other conclusions are more contentious, because they await the quantitative analyses needed to eliminate alternatives. Most would agree that the K–T boundary event occurred

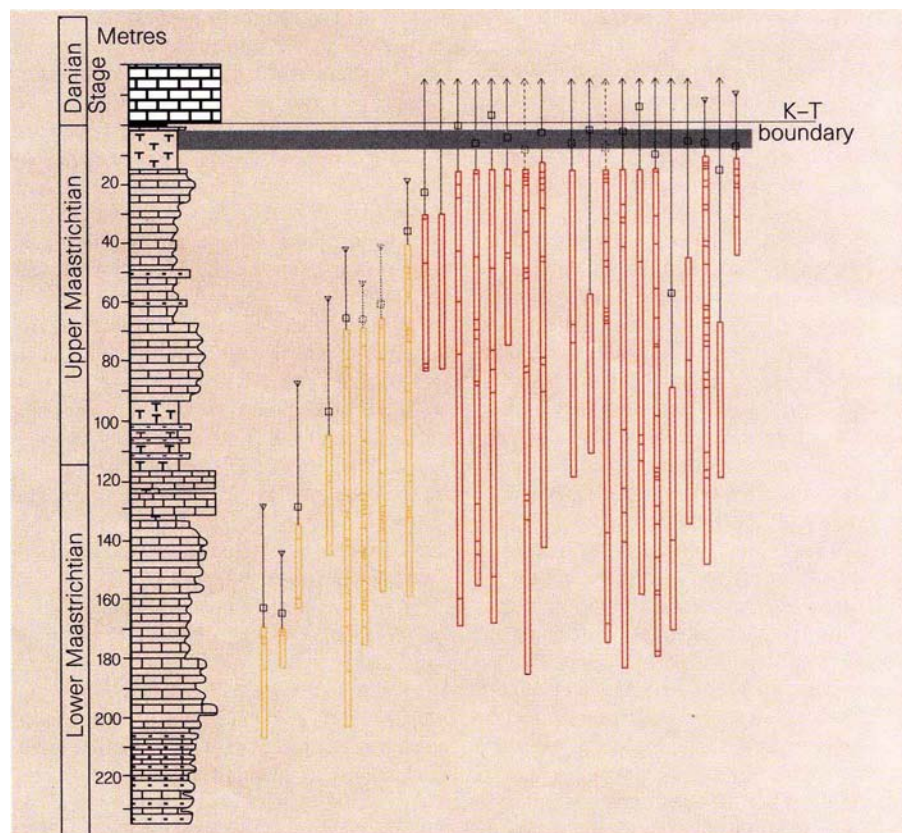


Figure 1 Confidence-limit analysis of extinction patterns of ammonite species immediately below the K–T boundary in the Bay of Biscay region. The ammonites seem to decline gradually, with the last in this dataset disappearing about 10 m below the K–T (Maastrichtian–Danian) boundary. However, most of the confidence limits calculated from collecting records (horizontal tick marks) extend into or above an unfossiliferous horizon (shaded) close to the K–T boundary. So most ammonite extinctions (those shown in red) are consistent with termination of the lineages at the K–T boundary itself rather than significantly below it. Boxes, 50 per cent confidence limits of range terminations; triangles, 95 per cent confidence limits. (Modified from ref. 17.)

within the context of longer-term changes in sea level, climate and other factors that have influenced biotic diversity throughout Earth history. Disentangling the effects of the sudden event from those of the larger trend is not simple, particularly when the environmental changes conspire to degrade the quality of the fossil record in the crucial interval. A drop in sea level, for example, will not only reduce habitable area in both shallow marine shelves and swampy coastal plains, and thus potentially drive extinctions; it will also reduce the area of outcrops or the breadth of palaeoenvironments available for sampling, and thus create potential biases that could artificially reduce observed biodiversity.

Much of the K–T debate has shifted to whether the apparent declines in diversity seen in groups as disparate as dinosaurs and ammonites during the last few million years before the K–T boundary are genuine biotic effects of such longer-term perturbations, or artefacts of sampling or preservation. For example, the apparent decline in global ammonite diversity through the last five million years or so of the Cretaceous — that is, from the Early to the Late Maastrichtian Stage — may reflect the drop in the number of known ammonite localities⁷. It could thus be a sampling artefact rather than a true biotic pattern, particularly because the diversity observed in each locality seems to hold steady over that interval.

Even without such biases, the natural patchiness of living populations and the incomplete nature of the fossil record mean that the last observed occurrence of a species rarely coincides with its final, true demise. So even abrupt extinctions will tend to look gradual or stepped, a phenomenon termed 'backwards smearing', or the Signor–Lipps effect^{4–8}. The rigour of palaeontological analysis has been greatly improved by the development of protocols, based on the number, placement and size of the gaps within the known temporal range of a species or evolutionary lineage, for putting confidence limits on the last observed occurrence in the rock record^{6,8}. This allows quantification of the intuitive perception that the extinction record of enormously abundant and resistant elements such as pollen and calcareous marine microfossils is more robust than that of rarer or poorly fossilized forms such as dinosaurs or sea cucumbers, and allows a new assessment of extinction patterns at critical localities and time intervals.

Turning again to the ammonites, for example, a literal reading of their record in some well-collected areas would be one of rapid but gradual decline leading up to the K–T boundary. But confidence-limit analyses on stratigraphical ranges in the classic sections at the Bay of Biscay⁹ and Seymour Island, Antarctica¹⁰, show that most of the observed range terminations are equally consistent with an abrupt extinc-

tion at the boundary (see Fig. 1).

As if these pitfalls were not enough, the churning activities of burrowing organisms or physical erosion and redeposition can produce a kind of 'forward smearing' of an abrupt event by artificially spreading the last occurrences of species beyond the true extinction horizon. This potential bias can now be factored out by using sediment-mixing models and time-specific isotopic signatures to recognize temporally displaced fossils^{5,11,12}; some species did indeed persist a bit above the K–T boundary, but many specimens have proven to be reworked.

This is not to say that all Late Cretaceous extinctions coincided with the K–T boundary. Confidence limits on the extinction of inoceramid bivalves, using the dissociated microstructural elements of the shell as distinctive microfossils and thus greatly increasing the number and density of records, show that this important group became extinct well before the end of the Cretaceous^{9,13}, and in a geographically complex pattern^{13,14}. It is still puzzling why so few other taxa were lost in whatever perturbation drove the disappearance of this once-pervasive and species-rich group.

The K–T debates are not over in the palaeontological community, but one gets the sense that they have turned a corner. The new methods that have been developed with the emergence of an increasingly rigorous quantitative palaeontology, collecting programmes designed in light of insights derived from palaeontological sampling theory^{8,15}, improved temporal correlations within the interval concerned¹⁶, greater interplay with other disciplines such as marine biology¹⁷ and geochemistry^{11,12} — all are combining to produce a clearer picture of one of the critical times in the history of life. □

David Jablonski is in the Department of Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA.

- Phillips, J. *Life on the Earth: Its Origin and Succession* (Macmillan, Cambridge, 1860).
- Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. *Science* **208**, 1095–1108 (1980).
- Sharpton, V. L. *et al. Geol. Soc. Am. Spec. Pap.* **307**, 55–74 (1996).
- Keller, G. *et al. Geol. Soc. Am. Bull.* **109**, 410–428 (1997).
- MacLeod, N. *et al. J. Geol. Soc. Lond.* **153**, 265–292 (1997).
- Jablonski, D. *Geol. Soc. Am. Spec. Pap.* **307**, 1–9 (1996).
- Ward, P. D. in *Ammonoid Biology* (eds Landman, N. H. *et al.*) 815–825 (Plenum, New York, 1996).
- Sepkoski, J. J. Jr & Koch, C. F. in *Global Events and Event Stratigraphy in the Phanerozoic* (ed. Walliser, O. H.) 21–34 (Springer, Berlin, 1996).
- Marshall, C. R. & Ward, P. D. *Science* **274**, 1360–1363 (1996).
- Marshall, C. R. *Geology* **23**, 731–734 (1995).
- Huber, B. T. *Geol. Soc. Am. Spec. Pap.* **307**, 319–334 (1996).
- MacLeod, K. G. & Huber, B. T. *Geology* **24**, 463–466 (1996).
- MacLeod, K. G., Huber, B. T. & Ward, P. D. *Geol. Soc. Am. Spec. Pap.* **307**, 361–373 (1996).
- Crame, J. A. & Luther, A. *Cret. Res.* **18**, 179–185 (1997).
- Marshall, C. R. in *The Adequacy of the Fossil Record* (eds Donovan, S. K. & Paul, C. R. C.) (Wiley, Chichester, in the press).
- Odin, G. S. *et al. Bull. Inst. R. Sci. Nat. Brussels, Sci. Terre* **66** (Suppl.), 111–117 (1996).
- Leviton, J. S. *Paleobiology* **22**, 104–112 (1996).

Daedalus

Forcible foam

Surface-active compounds, such as detergents and membrane lipids, have molecules in which a hydrophobic grouping is coupled to a hydrophilic one. They can stabilize a fluid interface by crowding along it as a monolayer, with the hydrophilic groups in the aqueous phase and the hydrophobic ones outside it. Certain biolipids have highly compact hydrophilic groups, but rather bulky hydrophobic ones. A monolayer of such wedge-like molecules therefore tends to impose a curvature on the interface, with the compact hydrophilic groups inside.

Daedalus is now extending this idea. His new foaming detergent has unusually bulky hydrophilic groups. Normally, aerated foams are unstable: small bubbles shrink while big ones expand and burst. But the molecules of the new surfactant will bend a surface into a concave curve, with gas inside it. It will stabilize small gas bubbles at a specific radius, defined by the packing geometry of its molecules.

Really stable gas-rich foams will have many uses. Fire-fighters will welcome them, brewers will use them to make even frothier beers, while manufacturers of cake-mixes and slimming breads will create new products of almost airborne lightness. Farmers may be able to grow their crops under a mulch of stable foam that lets light and air diffuse in, but keeps insects out. You might even be able to breathe such a foam, which would form an appealing environment for relaxation and play. But Daedalus's main goal is a foam to stabilize very small bubbles, maybe less than 50 nm across. Such tiny bubbles can exert up to a hundred atmospheres on the gas inside them. Daedalus's 'nanofoam' could hold as much gas, weight for weight, as a standard steel gas cylinder.

The obvious application is in rocketry. Existing rockets increase the density of their hydrogen and oxygen fuel by cooling them to liquids, with vastly expensive cryogenic consequences. Nanofoams of these gases could pack the fuels almost as densely, but could be stored and pumped at ambient temperature and pressure. A nanofoam would slowly lose gas by diffusion, but probably no faster than the evaporation of a cryogenic propellant.

A nanofoam could also be a high explosive. Butane nanofoamed in hydrogen peroxide, or oxygen nanofoamed in fuel oil, would be splendid blasting agents. And if the detonator failed, you need only wait a few hours or days for the gas to diffuse out, when the charge would be quite safe.

David Jones

George Wald (1906–97)

Biologist who discovered the role of vitamin A in vision

George Wald died on 12 April 1997, at his home in Cambridge, Massachusetts. He was 90. Wald's main scientific achievement was the discovery of the function of vitamin A in vision. Beginning with research dating to the early 1930s, he showed that the light-sensitive molecules, the visual pigments, found in the photoreceptor cells of the retina, consist of a protein (opsin) to which is bound a slightly oxidized form of vitamin A (vitamin A aldehyde, now termed retinal). Retinal serves as the chromophore for these molecules, absorbing the light and initiating conformational changes in the protein that lead eventually to the excitation of the photoreceptor cells. Wald's findings represented the first instance that a biochemical function for a fat-soluble vitamin had been established and were widely recognized.

Wald was born on 18 November 1906, in New York City. He grew up there and graduated from Washington Square College in 1927. He moved to Columbia University for graduate work, and studied there with Selig Hecht who was to introduce Wald to visual physiology. Hecht was the major figure in visual physiology of his generation; his quantitative studies of visual phenomena provided some of the first evidence that visual mechanisms can be explained in terms of physics and chemistry. Wald and his co-workers would eventually set many of Hecht's concepts into precise molecular terms.

Wald received his PhD degree in 1932 and spent the next year, his *Wanderjahr*, in Europe with three Nobel laureates — first with Otto Warburg in Berlin-Dahlem, then with Paul Karrer in Zürich and finally with Otto Myerhof in Heidelberg. It was during this momentous year that Wald identified vitamin A in the retina and gained the first glimpses of the part that it plays in vision. After a year at the University of Chicago he assumed his first academic position as tutor in biochemical sciences at Harvard. He remained at Harvard for all of his academic career, attaining full professorial status in 1948. He retired in 1977 as Higgins Professor of Biology.

Wald's research career centred almost exclusively on the chemistry of the visual pigments and related phenomena. Beyond his early fundamental discovery of the role of vitamin A in vision, he and his

colleagues made innumerable contributions to the biochemistry of the visual pigments. These included extensive studies on the chemistry of the rod pigment, rhodopsin, and the extraction and characterization of the first known cone pigment, iodopsin. His laboratory also elucidated the role of *cis-trans* isomerization in the visual process, demonstrating for the first time that such molecular transformations are involved in biological processes. He and his colleagues also provided insights into the diversity of visual pigments, vitamin A deficiency, visual adaptation, colour vision and the absorption properties of the cone pigments in primates, including man. No laboratory of that time added more to our understanding of the visual pigments and their relation to vision than did George Wald's. For his enormous contributions, Wald received the Nobel Prize in Physiology or Medicine in 1967, sharing the prize with the visual neurophysiologists H. Keffer Hartline and Ragnar Granit.

George Wald was a superb teacher, writer and lecturer. His interests ranged widely, and he had a special gift of drawing together diverse observations to make startling generalizations and proposals. Long before anyone had conceived of the notion that visual excitation might involve a biochemical cascade, he made such a proposal based on the then newly elucidated blood-clotting cascade. His studies on the distribution of the two forms of vitamin A found in nature (vitamin A₁ and A₂) led him into the field

of biochemical evolution, and he wrote superb and provocative articles on this topic with such varied titles as "The significance of vertebrate metamorphosis", "Life in the second and third periods; or why sulfur and phosphorus for high energy bonds?" and "The origin of optical activity". To his highly entertaining lecture "The origin of life", he added a sequel, "The origin of death"; both were enjoyed by audiences the world over.

Wald's courses at Harvard were exemplary for their wit and clarity. He taught introductory biochemistry to generations of Harvard students and, for 16 years, from 1960 until his retirement, a year-long introductory biology course entitled "The nature of living things". This course was part of the General Education Program at Harvard, taken both by students intending to concentrate on biology and by others. In 1966, *Time* magazine, in a cover story, named Wald "one of the ten best teachers in the country".

Wald was one of the first in academia to speak out against the Vietnam War, and in 1969 he gave a speech at the Massachusetts Institute of Technology that was to affect the rest of his life. The speech, "A generation in search of a future", immediately endeared him to the anti-war community and resulted in his becoming a forceful spokesman against the Vietnam War, nuclear arms proliferation and a variety of other political issues. In this, his wit and clarity of expression served him well. When confronted with the view that the arms race and the peacetime draft were "facts of life", he replied "No, those are the facts of death. I don't accept them and I advise you not to accept them". After his retirement in 1977, Wald gave up laboratory research and devoted his time to political causes. He travelled widely espousing his views until the last two years of his life.

When Wald received the Nobel prize in 1967, he remarked that "A scientist lives with all reality. There is nothing better. To know reality is to accept it and eventually to love it. A scientist is in a sense a learned child. There is something of the scientist in every child. Others must outgrow it. Scientists can stay that way all their lives". All who knew George Wald were profoundly affected by him. As Wald spoke of Hecht, his mentor, we speak of Wald: "He cast his light widely and many found their way by it".

John E. Dowling

John E. Dowling is in the Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Tyrannosaurs suffered from gout

Gout is a metabolic disorder in which urate crystals accumulate as space-occupying masses, producing monarticular spheroidal erosions in bone, often associated with new bone growth at their borders¹⁻³. We now report the first identification of such lesions in dinosaurs. Caricatures of the agony and ill-temper of those afflicted with gout are magnified by its recognition in *Tyrannosaurus rex*.

In a cast of the right forearm of the *T. rex* from Hell Creek Formation, South Dakota, popularly known as 'Sue'⁴ (specimen DMNH 30665; held at Denver Museum of Natural History), we observed that metacarpal I had a lesion of 11.5 × 9 mm with a slight rim (Fig. 1a). In addition, metacarpal II has a dorsal lesion of 7.1 × 5 mm, surrounded by an overhanging rim of bone (Fig. 1b), and a second medial surface erosion, 3 mm deep. This serendipitous observation led us to study other tyrannosaurids for erosive diseases.

Of 83 tyrannosaurid phalanges at the Royal Tyrrell Museum, only one specimen (TMP 92.36.328) has erosive lesions. The specimen is a partial tyrannosaurid pedal proximal phalanx (I-1) from Bonebed 149 (Upper Cretaceous), Dinosaur Provincial Park, Alberta, Canada. It has a defect (Fig. 1c) at the distal articular junction of subchondral and marginal bone. Slightly built-up bone forms an overhanging edge, overlying a smoothly excavated area of 4 × 9 mm, formed by the coalescence of two adjacent masses, as confirmed by radiological evaluation (Fig. 1d). There are no internal fronts or zones of bone resorption (Fig. 1) and there is no loss of perilesional bone density. Minimal filigree periosteal reaction is also present. Epi-illumination microscopy of the intact specimen, using polarizing optics, failed to reveal birefringent urate crystal preservation.

Gout is recognized clinically by documentation of urate crystal accumulation or by the recognition of characteristic radiological findings^{1-3,5}. Identification of the alkali-soluble crystals is usually not possible in archaeological, let alone palaeontological, specimens¹⁻³, and was not possible in this case. Thus, macroscopic and radiological appearance must form the basis for recognition of gout in prehistory. The uniformly excavated nature of the erosions in these specimens is characteristic of gout. Spheroid lesions with overhanging edges, common in gout, are only rarely reported in other diseases, such as multicentric reticulohistiocytosis, amyloidosis and type IIA hyperlipoproteinaemia¹⁻³.

The *Tyrannosaurus* erosions are quite distinct from the 'bite-like' erosions

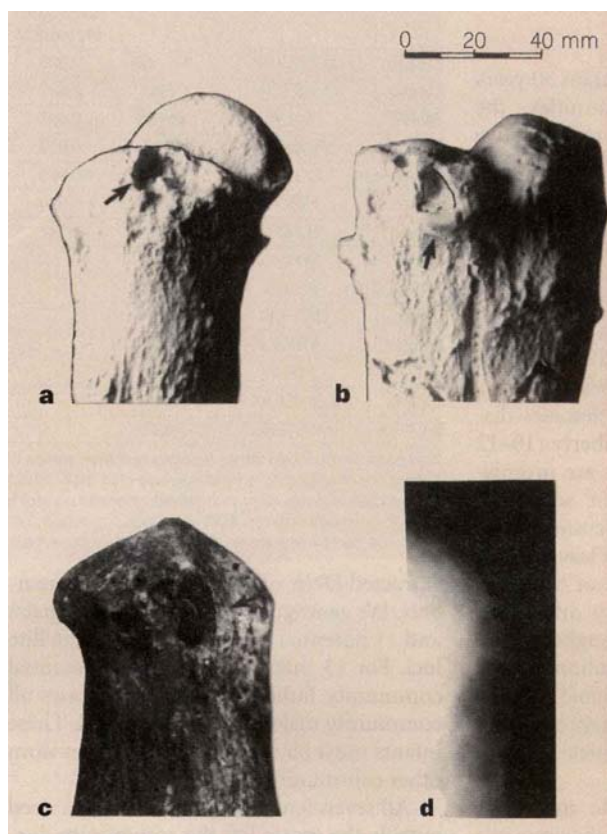


Figure 1 Dorsolateral views and X-ray of gouty lesions. **a**, DMNH 30665 metacarpal I. Erosion (arrow) with overhanging edge. **b**, DMNH 30665 metacarpal II. Spheroidal character of erosion is prominent at its base (delineated by a black line). Slightly overhanging edge (arrow) noted at lower margin of erosion. A second spheroidal erosion is present to the left and slightly above the first lesion. **c**, TMP 92.36.328 phalanx I-1. Oblong erosive process (arrow) consisting of two confluent spheroid erosions. Black area at the bottom of the erosion is an optical illusion. Arrow points to a slightly built-up adjacent bone. **d**, Oblique radiological view of TMP 92.36.328, illustrated in **c**. Two spheroidal erosions with overhanging edges.

characterized by the fronts of resorption seen in rheumatoid arthritis and spondyloarthropathy^{2,6}; the ill-defined lesions of calcium pyrophosphate deposition disease⁷; osteoarthritis (which does not produce bone erosions)¹⁻³; the draining sinuses and disorganized underlying osseous structure of infectious arthritis and osteomyelitis¹⁻³; and from osteosarcoma (which does not afflict joints)¹⁻³. A concurrent superficial bone infection in TMP 92.36.328 probably resulted from perforation of overlying skin by the gouty accumulation.

The apparent localization of erosion to the metacarpals and phalanges in tyrannosaurids may be the result of species variation. Only 15 per cent of human gout afflicts such joints, whereas first metatarsal phalangeal joints are more commonly affected (45 per cent of cases)^{1-3,5}.

Urate deposition in extant reptiles has been reported in both visceral and articular forms⁸. The latter has been reported in the monitor lizard (*Varanus*), turtles (*Testudo* and *Kinixys*), crocodilians (*Alligator* and *Crocodilus*) and in teguexin lizards (*Tupinambis*). These two occurrences in tyrannosaurids indicate that gout may have had a frequency similar to that observed in birds⁹.

Although dehydration (reported in reptiles)⁸ and renal failure (reported in birds)^{9,10} could be contributing factors in

Tyrannosaurus rex, a factor in humans is diet, by the ingestion of foods with a high purine content. One such dietary component is red meat, no stranger to this denizen from the Cretaceous. This tyrant king seems to have shared with subsequent tyrants the susceptibility to gout.

Bruce M. Rothschild

Arthritis Center of Northeast Ohio,
Youngstown,
Ohio 44512, USA

Darren Tanke

Royal Tyrrell Museum,
Drumheller,
Alberta T0Y 0Y0, Canada

Ken Carpenter

Denver Museum of Natural History,
Denver,
Colorado 80205, USA

1. Resnick, D. & Niwiyama, G. *Diagnosis of Bone and Joint Disorders* 2nd edn (Saunders, Philadelphia, 1989).
2. Rothschild, B. M. & Martin, L. *Paleopathology: Disease in the Fossil Record* (CRC, London, 1993).
3. Rothschild, B. M. & Heathcote, G. M. *Am. J. Phys. Anthropol.* **98**, 519-525 (1995).
4. Larsen, P. L. *J. Vert. Paleontol.* **11**, 41A-42A (1991).
5. McCarty, D. J. *Arthritis and Allied Conditions: A Textbook of Rheumatology* (Lea and Febiger, Philadelphia, 1989).
6. Rothschild, B. M. et al. *Am. J. Phys. Anthropol.* **82**, 441-449 (1990).
7. Rothschild, B. M., Woods, R. J. & Rothschild, C. *Clin. Exp. Rheumatol.* **10**, 557-564 (1992).
8. Appleby, E. C. & Siller, W. G. *J. Pathol. Bacteriol.* **80**, 427-430 (1960).
9. Schlumberger, H. G. *Lab. Invest.* **8**, 1304-1318 (1959).
10. Siller, W. G. *Lab. Invest.* **8**, 1319-1346 (1959).

Furtive mating in female chimpanzees

It has been assumed for more than 30 years that in chimpanzee communities the observed social unit and the reproductive unit are identical. We present genetic evidence that female chimpanzees seek to mate outside their own social group. We have studied a community of West African chimpanzees in which half the offspring were sired by non-community males.

Chimpanzees live in social groups of between 20 and 100 animals. Males remain in their community and defend home ranges of 7–30 km², whereas females disperse at around the time of puberty (10–12 years of age)^{1,2}. Most matings are promiscuous and opportunistic, but some are restricted by a dominant male and others may be exclusive, with couples leaving on a 'consortship' for several days or weeks^{1–3}. Inferring dates of conception is difficult as females are sexually active throughout their 15-day oestrus, and the gestation period has been shown to be variable⁴. Many questions about chimpanzee reproduction and female mating strategies remain unanswered.

Using non-invasive genetic methods⁵, we investigated reproductive behaviour in a community of West African *Pan troglodytes verus* from the Taï Forest, Ivory Coast (Fig. 1). The group has been studied for 17 years and has been fully habituated to human presence for over 10 years⁶. Both sexes interact continuously throughout most of the day, but at night each adult builds a tree-top nest and sleeps alone. The Taï community lives on a 25 km² territory surrounded by five non-habituated communities.

Between 1991 and 1995 we collected hair samples from nests and buccal-cell samples from chewed fruit, from which we

Table 1 Paternity of 14 infants in the Taï chimpanzee community

Offspring	Paternity				Behavioural data		
	Microsatellite genotypes		Paternity exclusion probability	Number of loci	Absence during oestrus (days)		Consortship with male
	Mother	Father			8 months before birth	7–9 months before birth	
LYCHEE	LOUKOUM	Macho	0.926	11	no data	no data	
Cacao	CASTOR	Fitz	0.887	9	15*	40*	Fitz
DORRY	DILLY	Kendo	0.967	10	1	3	
FEDORA	FOSSEY	Fitz	0.950	10	8	17	Fitz, Kendo
Gargantua	GOMA	Brutus	0.956	9	0	0	
Papot	PERLA	Rousseau	0.924	10	8	15	
VANILLE	VENUS	Ali	0.941	11	3	19	
PANDORA	PERLA	EGP		10	2	2	
BAGHEERA	BELLE	EGP		11	1	16	
Congo	CASTOR	EGP		11	2	11	
Hector	HERA	EGP		11	15*	45*	
HELENE	HERA	EGP		6	15*	45*	Macho, Brutus
Lefkas	LOUKOUM	EGP		8	1	1	
MOGNE	MYSTERE	EGP		10	5	26	Kendo

The name of the likely father (sharing common alleles at all 11 loci) from the same community is indicated. The probability of excluding a randomly sampled male from the same population using the same loci and allele frequencies is given. High probabilities indicate correct identification. Conception was estimated to have occurred 229±30 days before birth⁴. EGP, extra-group paternity. Females, upper case; males, lower case.

*Females completely absent throughout an oestrus were assigned an oestrus duration of 15 days per month.

extracted DNA of all 52 community members. We genotyped 21 mother–infant pairs and 11 potential fathers at 11 microsatellite loci. For 13 infants we tested all potential community fathers, and in seven cases all community males could be excluded. These infants must have been sired by males from other communities (Table 1).

All seven females who bore infants sired outside the group left the community during part of their oestrus period around the inferred time of conception (Table 1). This behaviour is surreptitious, and four of these seven were absent for only one or two days during the most probable conception period. Furthermore, during 17 years of observation we have never seen adult females approach neighbouring community males outside confrontational contexts. We witnessed several permanent transfers by foreign females to the study community but no surreptitious visits. As inter-community fights are the only opportunity for females to observe foreign males, the question arises as to whether females choose non-group mates on the basis of their performance during such encounters.

The females that have offspring with non-community males did not spend more time outside the group than females who conceived with group males. Thus, from the perspective of a male, a temporary absence from the group of a female during oestrus does not reliably indicate the extra-group paternity of group-born offspring. This is important as males may kill infants conceived outside their community².

Much significance has been attributed to consortships for male reproductive success^{1–3}. In this study, only two of six consortships resulted in paternity (Table 1). Also, a positive correlation between male

dominance rank and reproductive success in primates has often been suggested⁷. Small sample size precludes statistical tests of the association between male rank and reproductive success in the Taï Forest; however, four of the 11 potential fathers in our study were dominant males but only two of these sired infants during their tenure. Brutus and Macho (dominant for 10 and 1.5 years, respectively) sired no infants while dominant, but each sired one offspring afterwards. On the other hand, individuals that held dominant status for some time in their lives sired more than two-thirds of the offspring for which we identified likely fathers. Our results show that infants sired with neighbouring community females must be included when estimating male reproductive success.

Years of direct observation have supported the assumption that chimpanzees mate primarily within their social community, just as observations of supposedly monogamous birds did not reveal the significance of extra-pair copulations⁸. Possible cases of extra-group matings in chimpanzees were known¹ and in two cases (Gombe and Bossou)^{5,9} detected genetically, but the behaviour was dismissed as insignificant. Our data suggest otherwise, and underscore the need for comparable genetic surveys of other populations (now underway at Taï).

The significance of female-biased gene flow on chimpanzee social structure and evolution has been documented¹⁰. Our results show a further consequence of female behaviour, extra-group paternity, must now be taken into account when interpreting estimates of gene flow and genetic distance. That mate choice can occur within a larger area than the traditionally observed social units has implica-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Figure 1 Taï chimpanzees grooming.

tions not only for the evolution of mating systems and sociality, but also for the management of viable populations of these threatened primates.

Pascal Gagneux, David S. Woodruff

Department of Biology and Center
for Molecular Genetics,
University of California San Diego,
La Jolla, California 92093-0116, USA
e-mail: gagneux@jeeves.ucsd.edu

Christophe Boesch

Zoologisches Institut Universität Basel,
Rheinsprung 9, 4051 Basel, Switzerland

1. Goodall, J. *The Chimpanzees of Gombe: Patterns of Behavior* (Harvard Univ. Press, Cambridge, Massachusetts, 1986).
2. Nishida, T. (ed.) *The Chimpanzees of the Mahale Mountains* (Univ. Press of Tokyo, 1990).
3. Tutin, C. E. G. *Behav. Ecol. Sociobiol.* **6**, 29–38 (1979).
4. Martin, R. D. in *Paternity in Primates: Genetic Tests and Theories* (eds Martin, R. D., Dixon, A. F. & Wickings, E. J.) 238–274 (Karger, Basel, 1992).
5. Morin, P. A., Wallis, J., Moore, J. J. & Woodruff, D. S. *Molec. Ecol.* **3**, 469–478 (1994).
6. Boesch, C. in *Great Ape Societies* (eds McGrew, W. C., Marchant, L. F. & Nishida, T.) 101–113 (Cambridge Univ. Press, 1996).
7. Dunbar, R. I. M. & Colshaw, G. *Anim. Behav.* **44**, 1171–1173 (1992).
8. Birkhead, T. F. & Möller, P. *Sperm Competition in Birds* (Academic, New York, 1992).
9. Sugiyama, Y., Kawamoto, S., Takenaka, O., Kumazaki, K. & Miwa, N. *Primates* **34**, 545–552 (1993).
10. Morin, P. A. *et al. Science* **265**, 1193–1201 (1994).

Four climate cycles in Vostok ice core

The Russian–American and French international effort for drilling in ice achieved both a technical and a scientific success by reaching a depth of 3,350 m at the Russian Vostok station (78° S, 106° E; elevation 3,488 m; mean temperature –55 °C). In addition to being the deepest ice core, the Vostok core is now believed to cover the past four glacial–interglacial cycles (~400,000 years), a surprisingly long climate sequence which will be a valuable tool for palaeoclimatologists.

Hole number 5 started from the surface in 1990 and reached 2,755 m at the end of January 1994 (ref. 1). The station had to be closed for the winter because of financial and technical difficulties, but reopened the following summer season. Drilling operations were restarted and continued through the 1995 winter by three Russian drillers (headed by N. Vassiliev). By January 1996, the drilling had reached 3,350 m, but again was stopped because the station had to close for the winter.

The first 2,755-m-deep ice core had already provided the longest undisturbed ice climate record (Fig. 1). Studies of the deuterium (δD ; a proxy for temperature), dust and $\delta^{18}O$ records clearly showed that this core covers almost two full climate cycles^{1,2} (~240,000 years). The ice was processed during the 1995–96 field season for electrical conductivity measurements (ECM). Apart from volcanic events, the

background ECM signal is linked to acid concentration, and therefore depends on the balance between cation and anion concentrations. This changes dramatically between glacial and interglacial times as a result of changes in the intensity in sources of marine and continental aerosols and/or changes in their transport by atmospheric circulation^{3–5}. High conductivity (acidity) is observed for the last interglacial ice (marine stage 5.5) as well as for the three warm sub-stages of stage 7, whereas the signal appears lower for stage 6 and cold substages of stage 7. Despite noticeable differences, ECM gives a stratigraphy consistent with that provided by the isotope record.

Below 2,755 m, we observe two long intervals of sustained high electrical conductivity, at 3,073–3,124 m and 3,262–3,298 m depth in the core. We believe that these correspond to interglacial stages 9 and 11, respectively. This is confirmed by deuterium data as well as by preliminary dust measurements, which both fall between Holocene and last interglacial levels. An ice layer at 3,020–3,035 m has a similar ‘interglacial’ ECM signal but no clear deuterium signature in stable isotopes, possibly because of the insufficient time resolution at this period. Warm marine substages 9.1 and 9.3 may correspond to intervals 3,020–3,035 m and 3,073–3,124 m, respectively. For ice

deeper than 3,124 m, the ECM fine structure (with patterns also depending on cation and anion concentrations; I. B., manuscript in preparation) suggests that the high dust input of glacial stage 10 started at 3,139 m. This level, dated to 340,000 years in the SPECMAP chronology⁶, should mark the boundary between stages 9.3 and 10. The interval 3,139–3,250 m probably corresponds to cold stage 10 and the high, short ECM event around 3,180 m results from a high aerosol concentration. Before warm stage 11 (below 3,298 m), the deep ice close to 3,350 m displays a low ECM signal, low deuterium content and high dust concentrations. This deepest part was thus probably deposited at the end of the previous glacial period (that is, at stage 12).

For a very preliminary age estimate, we assume that the ice layers thin linearly with depth below 2,755 m (240,000 years ago) and adopt an age of 340,000 years at 3,139 m depth. Although simple, this approach appears reasonable: age at 3,030 m (310,000 years) is consistent with stage 9.1; ages at 3,262 and 3,298 m (385,000 and 400,000 years) are more or less in agreement with stage 11, and 426,000 years at 3,350 m is consistent with the end of stage 12.

A wealth of climate-related information will be soon available back to 400,000 years or so ago. The period from 240,000–400,000 years is recorded in about 600 m of ice, which will allow us to study it with high resolution: 1 m of ice would represent 200 years at 2,755 m, 260 at 3,000 m and 450 at 3,300 m. Current measurements include the crystalline texture of the ice, the isotopic composition and concentration of aerosols, chemical constituents and cosmogenic isotopes, as well as the elemental and isotopic composition of the entrapped air bubbles. The last 300 m of ice to be drilled (drilling will be stopped 50 m above the water lake⁷) may have the potential to extend the climate sequence still further back in time, although disturbances of the layers, as seen for the deep ice in the Greenland ice core, are likely to appear.

J. R. Petit, I. Basile, A. Leruyet

D. Raynaud, C. Lorius

LGGE, CNRS, BP 96,

38402 St Martin d'Hères Cedex, France

e-mail: petit@glaciog.ujf-grenoble.fr

J. Jouzel, M. Stievenard

LMCE, CEA/DSM CEA Saclay,

91191 Gif-sur-Yvette, France

V. Y. Lipenkov, N. I. Barkov

AARI, Beringa Street 38, 199226, St Petersburg, Russia

B. B. Kudryashov

SMPI, 2-22 Linya, 199026, St Petersburg, Russia

M. Davis, E. Saltzman

RSMAS, 4600 Rickenbacker Causeway,

Miami, Florida 33149-1098, USA

V. Kotlyakov

IGAN, Staromonetmy, per 29,

109017 Moscow, Russia

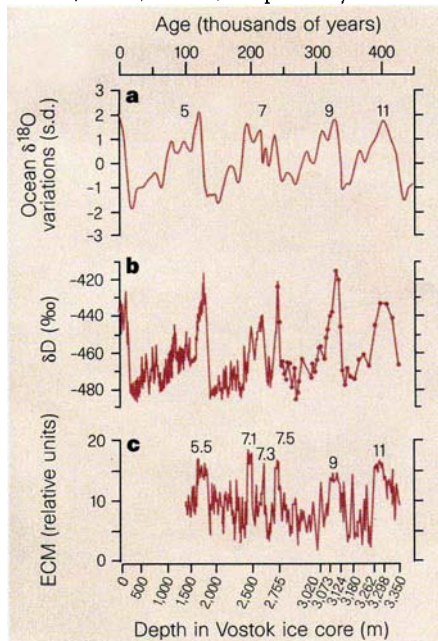


Figure 1 Vostok and marine climate records over the past 400,000 years. **a**, Ocean $\delta^{18}O$ variations (standard deviation units) deduced from foraminifera data from deep-sea cores (SPECMAP stacked record from Imbrie *et al.*⁶). **b**, Vostok δD values (expressed as ‰ with respect to Standard Mean Ocean Water). The continuous deuterium profile down to 2,755 m is from Jouzel *et al.*¹ and from a discontinuous set of samples below this depth. **c**, Vostok ECM signal (3-m running mean and expressed in relative units) for depth below 1,500 m. Numbers indicate marine stages.

1. Jouzel, J. *et al. Clim. Dyn.* 12, 513–521 (1996).
2. Waelbroeck, C. *et al. Clim. Dyn.* 12, 113–123 (1995).
3. De Angelis, M., Barkov, N. I. & Petrov, V. N. *Nature* 325, 318–321 (1987).
4. Legrand, M., Lorius, C., Barkov, N. I. & Petrov, V. N. *Atmos. Environ.* 22, 317–331 (1988).
5. Petit, J. R. *et al. Nature* 343, 56–58 (1990).
6. Imbrie, J. *et al. in Milankovitch and Climate Pt 1* (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (Riedel, Hingham, MA, 1984).
7. Kapitzka, A. P., Ridley, J. K., Robin, G. d. Q., Siegert, M. J. & Zotikov, I. A. *Nature* 381, 684–686 (1996).

Evidence for stone age cranial surgery

In 1996 an exceptionally well preserved skeleton was excavated at the stone age burial site of Ensisheim (Alsace). The cranium of the buried individual shows clear evidence of two trepanations (surgically created holes in the skull). Signs of long-term healing indicate that this type of *intra vitam* surgical intervention was skilfully practised more than 7,000 years ago. Our findings from the Ensisheim skeleton represent the earliest unequivocal evidence of healed trepanations yet discovered.

The Neolithic burial site of Ensisheim is among the best documented necropolises of early agriculturalists in France¹. Its archaeological context is representative of younger linear pottery (5200–4900 BC) in central Europe. Burial number 44, discovered in September 1996, contained the well-preserved remains of a man who died at roughly 50 years of age. Grave goods assign

a date of 5100–4900 BC typologically and a ¹⁴C-estimation of the age of the human bones validates the archaeological data (Utrecht ¹⁴C laboratory sample UtC-5406: 155±39 radiocarbon years before present, ~5100 BC).

Trepanation is the surgical removal of sections of the cranial vault during life². The skull from burial 44 has two of these surgical openings, both of which show clear evidence of healing (Fig. 1). The anterior opening shows solid and complete osseous restoration of the defect by bony healing. The posterior opening shows only partial healing. Restored surfaces are quite thin in places, probably due to the enormous size of the defect. As a rule, for osseous defects larger than 5 cm in diameter, residual openings are left in modern clinical cases^{3,4}. If complications such as haemorrhage, brain damage, wound infection or meningitis do not occur after craniotomy, and if primary bone healing takes place, reactive processes fail to develop and long-term survival is observable^{5,6}.

Holes in ancient skulls may be caused by infections, tumours, fractures, post-mortem animal activities, damage during disinterment or by selective erosion^{2,6}. All of these defects can resemble trepanation openings on superficial investigation. In the present case, diagnosis is unequivocal. Our findings clearly indicate that the individual underwent *intra vitam* surgical treatment twice, either consecutively or simultaneously.

Field studies on trepanation provide valuable insights for reconstructing the

motives for such neurosurgery⁷. In native African communities, there are traditionally two motives for trepanations: therapeutic, for head injuries such as fractures; or 'magical'/spiritual therapy, for head (brain) disorders such as persistent headaches, epilepsy, intracranial tumours, and mental disease. Motives inferred for the execution of trepanations in prehistoric times must necessarily remain speculative. In the man from burial 44, we could not find any indications of ailments requiring therapeutic measures, but their existence cannot be ruled out.

Trepanations are among the most fascinating surgical operations in human history executed in both the Old and New Worlds^{8,9}. Up to now, claims to cases pre-dating the Late Neolithic age^{10,11} have been highly dubious. The case we describe is exceptional for several reasons: it appears to be the oldest healed neurosurgical operation known worldwide, its technical realization testifies to the high craftsmanship and well-founded anatomical knowledge of the surgeon, and the success of the unusual trepanations is established by the long survival of the 'patient'.

Kurt W. Alt

*Institute of Human Genetics and Anthropology, Freiburg University, Breisacher Strasse 33, D-79106 Freiburg, Germany
e-mail: alt@uni-freiburg.de*

Christian Jeunesse

Service Régionale de l'Archéologie d'Alsace, Palais du Rhin, 2 Place de la République, F-67082 Strasbourg, France

Carlos H. Buitrago-Téllez

Rüdiger Wächter*

*Department of Diagnostic Radiology, *Department of Craniomaxillofacial Surgery, Freiburg University, Hugstetter Strasse 55, D-79106 Freiburg, Germany*

Eric Boës

Laboratoire d'Anthropologie, University of Bordeaux, Avenue de Faculté, F-33405 Talence Cedex, Bordeaux, France

Sandra L. Pichler

Department of Prehistory, Freiburg University, Belfortstrasse 21, D-79085 Freiburg, Germany

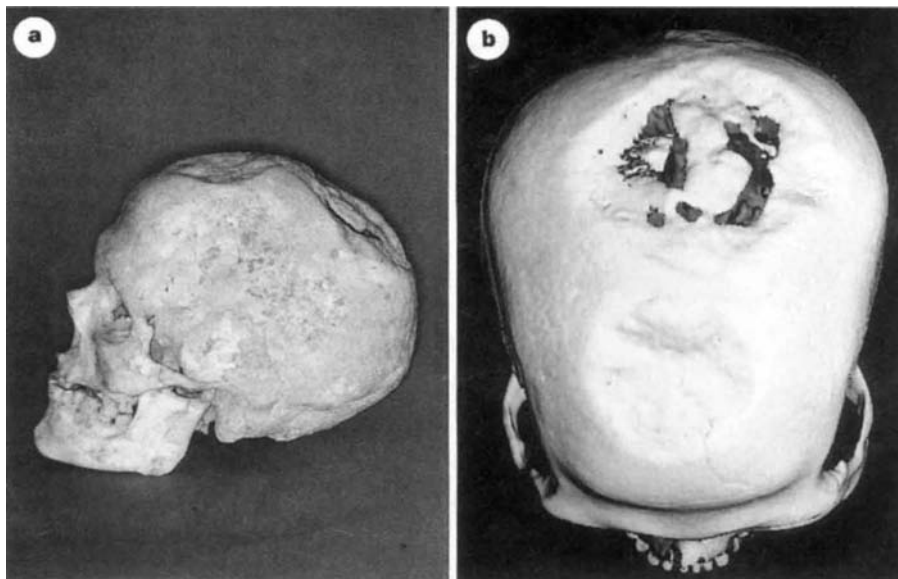


Figure 1 **a**, Photograph of the cranium, lateral view. **b**, Three-dimensional computed-tomography reconstruction of the cranium, superior view (details of the technical procedures are available on request from the authors). The two trepanations are quite similar in shape. The anterior, smaller trepanation (6.5×6.0 cm) is located in the frontal bone, nearly reaching the coronal suture, and extends slightly more to the right side of the cranium. The orifice is closed completely by advanced reactive bone development. The larger, rhombic trepanation (9.5×9.0 cm) extends over both parietal bones, with larger portions on the left side of the cranium. Two- and three-dimensional computed-tomography reconstructions of the cranium show complete consolidation at the borders of the trepanation with increased sclerosis.

1. Jeunesse, C., Lambach, F., Mathieu, G. & Mauvilly, M. *Cah. Ass. Promot. Rech. Archeol. Alsace* 9, 81–88 (1993).
2. Lisowski, F. P. in *Disease in Antiquity* (eds Brothwell, D. R. & Sandison, A. T.) 651–672 (Thomas, Springfield, 1967).
3. Persson, K. M., Roy, W. A., Persing, J. A., Rodeheaver, G. T. & Winn, H. R. *J. Neurosurg.* 50, 187–197 (1979).
4. David, D. J., Poswillo, D. & Simpson, D. *The Craniotomies* (Springer, Berlin, 1982).
5. Horwitz, N. H. & Rizzoli, H. V. *Postoperative Complications of Intracranial Neurological Surgery* (Williams & Wilkins, Baltimore, 1986).
6. Pahl, W. M. *Ägyptische Schädelchirurgie* (Fischer, Stuttgart, 1993).
7. Margreth, E. L. in *Disease in Antiquity* (eds Brothwell, D. R. & Sandison, A. T.) 673–701 (Thomas, Springfield, 1967).
8. Bakay, L. *An Early History of Craniotomy* (Thomas, Springfield, 1985).
9. Rogers, S. L. *Primitive Surgery* (Thomas, Springfield, 1985).
10. Piggott, S. *Proc. Prehist. Soc.* 6, 112–133 (1940).
11. Ullrich, H. & Weickmann, F. *Anthrop. Anz.* 29, 261–272 (1965).

Future shocks and present woes

Imagined Worlds

by Freeman Dyson

Harvard University Press: 1997. Pp. 216. \$22, £14.50

Oliver Morton

If you were to ask so-called 'hard' science fiction writers — the ones who try to maximize the boggling of minds while minimizing the flouting of known science — to name their favourite scientist, Freeman Dyson would stand a good chance of coming top. It's easy to see why. Dyson has a startlingly profound imagination, a willingness to take ideas as far as they can possibly go.

He suggests that truly advanced civilizations will one day encapsulate whole suns so as not to waste energy; and he worries about whether there can be enough unanswered questions in the world to keep immortals interested. In this book he provides a fascinatingly plausible view of artificial telepathy. He has helped to design extraordinary spaceships and advised the Pentagon on wild (and no doubt occasionally woolly) weapons. Best of all, from the science-fiction writer's point of view, he admires science-fiction writers. This book is, in part, a tribute to science-fiction; it is an attempt not to predict the future, but rather, through imagination, to bring some of its potential to life.

Like Dyson's earlier book, *Infinite in All Directions* (Harper and Row, 1988), *Imagined Worlds* is based on a series of lectures, in this case given at the Hebrew University of Jerusalem in 1995. At its heart are three prophetic chapters, one on science, one on technology, and one on both of the above and more-or-less everything else. As a curtain-raiser he offers a set of anecdotal arguments as to why it is best to let the future work itself out through trial and error, rather than trying to force its pace for any ideological reason. And as a tail-piece he considers some of the ethical issues his futures open up.

Those who have enjoyed Dyson's previous books will find much of this one familiar — happily so in the elegance of its expression and the range of its interest. Although the examples and targets are different, the ideas illustrated and attacked are the same as they have been since he started writing. The preference for trying lots of small things rather than a few big things is vintage Dyson (and will not be very popular with those of his Princeton University neighbours who are fierce advocates of vast tokamak fusion reactors).

His belief in biotechnology as something much more than a source of medical advances — as a source, for example, of pet dinosaurs, their genetic blueprints worked out from scratch rather than recovered from

amber, of factories that eat their own effluent, of potatoes designed to grow on Mars — is another old favourite. So is the vision of human life spreading throughout the planetary system, into the Kuiper belt, the Oort cloud and eventually to other stars. This is another version of Dyson's 'small and many is best' philosophy, one in which people with different ideas of what it might be to be human, or post-human, simply put a few light hours between their cometary homes and agree to differ.

There are few other writers who would allow themselves to cover such a range. In Dyson's youth there were rather more: H. G. Wells and his heirs, J. B. S. Haldane, Desmond Bernal, Olaf Stapledon. They were writers on science who, having looked into the abyss in the First World War, were willing to look into the depths of the future with strange mixtures of faith and pessimism. Dyson is the heir not just to their range, but also to some of their pessimism — and it is that pessimism which is responsible for the book's major flaw, its treatment of the present.

Dyson quotes with approbation G. H. Hardy's view that "a science is said to be useful if its development tends to accentuate the existing distribution of wealth, or more

directly promotes the destruction of human life". He similarly subscribes to Haldane's claim that "the tendency of applied science is to magnify injustices until they become too intolerable to be borne". These are strong claims, more shocking to most of Dyson's peers, I would imagine, than they would have been in the Cambridge of the 1920s and 1930s; they require considerable justification if we are to take them seriously. And that justification does not arrive.

Dyson sees the parts of the world he has lived in as being in serious decline. "In the 50 years since [Wells] died, England has gradually reverted to a class system with inequalities almost as sharp as those that he fought against as a young man and lampooned in his novels." The United States, too, is going downhill, and Dyson believes that the social disintegration he has observed since moving there 50 years ago is largely due to the effects of technological change derived from scientific progress: "Because of science, children deprived of legitimate opportunities to earn a living have strong economic incentives to join gangs and become criminals."

There are two problems here. One is that some of these observations are just wrong. The United Kingdom is not as equal a society



Hats off to the bike

The bicycle has been an enduring technological success story. World bicycle production has been on a steadily upward trend for most of the postwar period. In China (above) bikes are still

the most common delivery vehicles. The social and technical history of the subject is explored in the profusely illustrated book, *The Bicycle*, by Pryor Dodge (Flammarion, £35, \$50).

as some of its European neighbours, but it has more university graduates than ever before and, according to the polls, it has more republicans than Tories. Saying that it has regressed to the standards of a highly stratified age in which much of the working-class majority lived in extremely poor conditions is not only silly, it also makes it harder to recognize the patterns of real hardship and despair that persist among an excluded minority.

The second problem is ascribing these lamentable new patterns of destitution to science. There is no doubt that technology derived from science has changed the distribution of jobs in the United States, and that there are real costs to this transition, costs that hurt people. But the United States continues to produce new jobs at what seems from the other side of the Atlantic to be an extraordinary rate, and it is simply not the case that all this new employment is in the form of worthless McJobs. Many are good jobs that depend on technology.

The fact that some people are excluded is a real and terrible problem. But to suggest that it is a natural outcome of the application of science rather than being caused by "drugs, or by guns, or by racial intolerance, or by bad schools, or by broken families" just seems strange. If the problems of failed prohibition, epidemic homicide, ingrained prejudice, chronically under-achieving education and parental absence or ineffectiveness could be solved, does Dyson really think that the 'underlying' problems of technological innovation in the world of work would still drive things downhill? Does anyone else?

The sad aspect of all this is that the world desperately needs people with the insight Dyson is capable of to look searchingly at some of these problems; it needs wise men who know that there is more to life than economic growth. In his introduction, Dyson talks of the need to look at the future without a seven per-cent discount rate, to look outside the eternal now of the evermore-frictionless markets championed by his daughter Esther, business guru by appointment to the wired cognoscenti. But he does not really deliver on this excellent beginning; he talks about decline without spelling out what to do about it, and he talks about moral dilemmas without clarifying them to any great extent.

Yet he does not give the impression that he is not up to the task — rather he seems to get side-tracked, and not to bother to criticize some of his own assumptions. In short, I think Dyson could write a better book about the future than this one. I think he could write a better book about the future than almost anyone; I wish he would. □

Oliver Morton, formerly science and technology editor of *The Economist*, is a freelance writer and contributing editor at *Wired* and *Newsweek International*.

Deciphering an enigma

A Logical Journey: From Gödel to Philosophy (Representation of Mind)

by Hao Wang

MIT Press: 1996. Pp. 402. \$40, £33.95

Logical Dilemmas: The Life and Death of Kurt Gödel

by John Dawson

A. K. Peters: 1996. Pp. 376. \$49.95, £36

Karl Sigmund

Late in life, the mathematician Kurt Gödel confided that he felt the Institute for Advanced Study at Princeton University had been paying him too much. On the face of it, he had a case: after his appointment as professor, he never lectured or held a seminar, and he published almost nothing. But the man who, according to the citation of his honorary Harvard degree, had "discovered the most significant mathematical result of this century" — the algorithmical inexhaustibility of mathematics — was far from idle during his years of outward tranquillity at Princeton. He struggled with a huge philosophical task, convinced that it was possible to "do for metaphysics as much as Newton did for physics".

No love was lost between Ludwig Wittgenstein and Gödel — Hao Wang writes that "Gödel's habitual calmness was absent" when he spoke of his compatriot — but the biographies of these two thinkers show striking parallels. Both men achieved a towering reputation through one short piece of work done while still in their twenties; both subsequently produced a huge pile of written material which they kept increasing and restructuring without finding a final form; both closely supervised the efforts of disciples to present their thoughts vicariously (Friedrich Waismann for Wittgenstein and Wang for Gödel); and both left a vast written legacy, fuelling marathon projects of continuing collected works published posthumously.

Wittgenstein's legacy is the larger, but Gödel's has the extra touch of being written in *Gabelsberger*, a shorthand that is now defunct. Gödel, who was nothing if not circumspect, knew that it could serve as a cipher: at Notre Dame University, he had recorded the text of his lectures in longhand and encoded the examination questions in *Gabelsberger*. It took John Dawson two years to catalogue the 60 boxes containing Gödel's scientific legacy. He was greatly helped in this task by his wife Cheryl, who learned *Gabelsberger*.

Dawson and Wang have opened two breaches into the wall of self-imposed silence surrounding Gödel's thoughts. Dawson's account of Gödel's life and work is a masterpiece. This carefully researched and eminent-



Gödel: out of step with his time.

ly readable account covers Gödel's protected childhood, his early intellectual triumphs and psychic crises in Vienna, his dramatic flight around a war-torn globe to a secure existence at Princeton and mothered by a devoted wife, the growing recognition of his work, his gradual withdrawal and the poignant deterioration leading to death by self-starvation. Gödel's historic achievements are described with skill and precision. These include the famous incompleteness theorems, the consistency of the continuum hypothesis with the axioms of set theory, and the existence of closed time-like solutions for Einstein's field equations.

Despite some tantalizing gaps, there is no dearth of biographical material about this intensely secretive man. He kept every piece of paper — including stubs of cinema tickets, drafts of unsent letters, and slips of library requests — and wrote to his mother every month for several decades. Solemn and uptight as he was, Gödel had distinguished friends, including Einstein, John von Neumann, and his fellow-exiles from Vienna, the mathematician Karl Menger, the economist Oskar Morgenstern and the philosopher Rudolf Carnap, who left extensive records ably put to use in Dawson's book. If Gödel's figure remains strangely impersonal, it is because that was how he was.

Morgenstern wrote that talking to Gödel was like "being thrust immediately into another world" — a world that was utterly self-contained, somehow like a monad as conceived by Gödel's hero Leibniz. Everything from outside was met with massive distrust. Gödel usually felt out of step with his time. He rarely agreed with Einstein, for instance. He vehemently opposed Wittgenstein's language analysis, and the logical positivism of his teachers from the Vienna Circle. Gödel totally upset the debate on the foundations in mathematics, waged between Hilbert's formalists and Brouwer's intuitionists, by showing that a finitary system allows no proof of consistency within

the system, and that intuitionist mathematics is no safer from contradiction than classical mathematics.

The implications drawn by Gödel from his own results were also quite opposed to everyone else's. He defended his unadulterated platonism by claiming that it was both a prerequisite and a consequence of his discovery of the incompleteness theorems. A firm conviction that concepts do exist objectively may indeed have given him the confidence he needed for his labours (most mathematicians seem to be platonists during their working hours). But using an eventual reward as argument for the truth of a persuasion smacks of theology.

Gödel's other thesis had also a scholastic ring to it. He claimed that "the creator necessarily knows all properties of his creatures"; as mathematics is mechanically inexhaustible, it follows, roughly speaking, that the mind is no machine, or mathematics no creation of the mind, or both. Each alternative implies the existence of a non-physical reality. Gödel pursued this line of argument, for instance, in his "Carnap paper" which he kept rewriting from 1953 to 1959, producing six versions and driving his editor to exasperation. Although he shied away from committing himself in print, the conviction of the reality of abstract entities was so much part of him that his own existence became increasingly abstract and disembodied.

Dawson notes that Alan Turing has been made the hero of a play, and speculates whether a playwright will dare to dramatize Gödel's life. Samuel Beckett's *Endgame* comes to mind. All the material needed for the ultimate monologue can be found in Wang's *A Logical Journey*. This is the third volume based on Wang's conversations with the Princeton recluse (the others are *Reflections on Kurt Gödel* and *From Mathematics to Philosophy*).

The interviews took place in the 1970s, at a time when Gödel's statements were often rather cryptic. Wang sifts through them again and again, like an archaeologist. He patiently attempts to reconstruct eerie allusions on what the philosopher Husserl did not write down; on secret but fruitful philosophies in Russia; or on life force as a primitive element in the Universe. And Wang politely compares "the persuasiveness of Gödel's more or less finished writings with the unreasonableness of his speculations".

Sadly, Wang died before his book went to press. The editors suggest that the author would have removed some of the repetitions, but this difficult book needs no editing. It documents the tragic struggle of Gödel's final years, and resembles a silent wreck set adrift after having been ice-bound for years. This book is a monument to a discoverer who had ventured further than anyone else. □

Karl Sigmund is at the Institute of Mathematics, University of Vienna, Strudlhofgasse, A-1090, Austria.

Sex, flies and videotape

Microcosmos

A film by Claude Nuridsany and Marie Pérennou

Galatee Films/Jacques Perrin: 1996. Various cinemas in Europe and the United States

Helen Phillips

Microcosmos has to be one of the most unlikely film successes of recent years and is now trying its luck with UK audiences. True, it has all the elements one might expect from the French film industry: romance, sex, drama and humour, set against a background of beautiful scenery. But there are no 'actors' — just insects (plus the odd spider and snail) doing what they do. The story is simply one day in the life of the tiny inhabitants of a field, somewhere in rural France.

The scale of large-screen cinema aside, this is no ordinary wildlife documentary. The directors, Claude Nuridsany and Marie Pérennou, reject traditional academic life to concentrate on less conventional ways of communicating their research and, perhaps more importantly, their passion for all creatures small and... small. In teaming up with the actor and producer Jacques Perrin, they have distilled years of observation into a 75-minute feature film, producing a fantasia of portraits of life in miniature.

We are never explicitly told anything about the creatures; not even their names, Latin or common — the subjects are simply credited as the "cast" at the end of the film. There is no commentary apart from a brief introduction (by Kristin Scott Thomas in the English translation). We are

asked to "fall silent and listen", and this is exactly what we do.

The early part of the film is a little like an animated coffee-table book, with insects engrossed in their morning grooming, a butterfly emerging from its chrysalis, and a dance of time-lapse images as waving tendrils echo the movement of the looper caterpillar. A strange bee's-eye-view computer graphic is rather incongruous, but as the creatures move off in search of food the picture book gives way to a series of tales that would soften the heart even of those who would not think twice before swatting a fly.

The film-makers tried to film insects with 'personalities', to show that they are not so unlike us. This anthropomorphic aim is rather off-putting, but it is only really reflected in the choice of subjects and scenes. The directors aim to avoid portraying the creatures as automatons, and attempt to place violence in context. They do succeed in showing insects as highly alert — antennae twitch, caterpillars track a bee buzzing overhead, and stag beetles eye each other before launching into battle. It is best to forget the cute aims and just enjoy it.

Heroes of the film include a persistent and comical dung beetle, battling to free his ball of dung from a spine cruelly placed in his path for our entertainment, and a snake-like train of furry caterpillars, demonstrating traffic control to be envied. Even the hoarding ants are somehow much more appealing at this large scale. An erotic slimy dance of two mating snails is filmed (without garlic butter) in a truly romantic way. There is more romance as a bee tries to mate with an orchid mimic, only to be rewarded with a pollen parcel. Drama comes in the form of a rainstorm, frightening at this scale, catapulting ladybirds from



Mantis larva praying for a big hit with macrocinematic extravaganza *Microcosmos*.

their perches and flattening the odd ant.

The technical achievement in producing this film can only be applauded. Planning and design took two years, filming another three, during which time the team developed its 'macrocinematic' techniques, with special cameras, radio-controlled camera mounts, lighting set-ups that did not bake the cast and, perhaps most difficult of all, suitable conditions to keep the insects happy and performing despite being only an antenna-length from the lens.

The soundtrack is superb. Music by Bruno Coulais is merged with sounds collected on location. The creatures' clattering footsteps, chirping and buzzing are truly impressive at this scale. Bee-flies sound like a motor racing team, and a millipede drifts along like a steam train.

The words of the final song — "open your eyes before you die" to this "world beyond imagination yet beneath our notice" (sung by an angelic choirboy) — makes one wish that even the small amount of commentary had not been translated from the French. But one has to admire the skill and beauty of the film. And when things threaten to get just a little too romantic, something else (and I almost said someone else) gets eaten.

Helen Phillips is on the editorial staff of Nature.

New journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 11 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals that first appeared during or after June 1995 and issued at least four separate numbers by the end of May 1997 will be considered.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are publications of abstracts.
- Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.
- Deadline for submission is 6 June. When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates. For further information please contact Peter Tallack, Book Reviews Editor, *Nature*, Macmillan Magazines, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

In retrospect chosen by David Hughes

Frontiers of Astronomy

by Fred Hoyle
(1955)

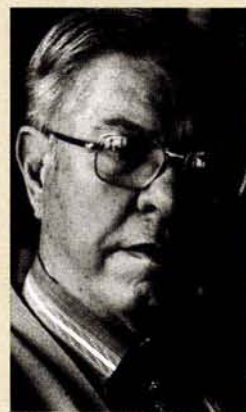
Most of today's astronomers experienced, when young, a 'road to Damascus' conversion that seduced them from the school-taught subjects such as physics and mathematics, and diverted their paths towards astronomy. Hoyle has been responsible for many of these conversions, and his *Frontiers of Astronomy* was often the 'holy book'. Here Hoyle is happy to divide scientists into two groups, these being astronomers and the rest, and he makes absolutely no bones about telling the world which category he thinks is the superior.

Hoyle stresses the fact that the joy of being an astronomer is that in many ways you are helpless. The Universe cannot be touched. Whereas other scientists do experiments, astronomers merely observe. Astronomers cannot go on 'field trips' and wander off around the Galaxy investigating objects of interest. An astronomer cannot tear to pieces objects such as planets and stars, using techniques much loved by physicists. Even though astronomers can build big telescopes and efficient detectors, in essence they are meek. They cannot alter the radiation that comes into the telescope, or experiment with its source. They have to sit here on Earth and simply accept what comes down the tube.

But they do have one overwhelming advantage, and that is variety. The Universe is so vast, and the expanses of time that are of astronomical interest are so great, that almost every conceivable type of astronomical process is still taking place somewhere. In many astronomical fields there is an embarrassingly huge flood of information. In others the astronomers are convinced that if only they could look a little further, or probe back in time slightly more, or slightly increase the resolution of their images and spectra, the picture would become clear. And it requires the straitjacket of astronomical theory not only to untangle and decipher the information flood but also to encourage the desired instrumental advances.

Hoyle also divides astronomers into two types. On the one hand there are those who concentrate solely on the known. They give the impression that the job of cataloguing, exploring and understanding the Universe is nearly complete. But Hoyle will have nothing to do with this complacency. Hoyle's astronomers are pioneers; they are the vanguard troops in the scientific campaign against the unknown.

This book concentrates on the oddities and mysteries. And they are not just the astronomical mysteries that, by chance, occupied the mind of a Cambridge lecturer in mathematics, in his late thirties, writing in 1954, three years before the dawn of the space age. It is amazing just how many of Hoyle's mysteries are still unsolved today. The range can be judged by the questions he



Hoyle: mystery writer.

poses. Close to home, he worries about why earthquakes do not occur at depths below 700 km and what the temperature of the Atlantic Ocean was 200 million years ago. He then ponders why the Sun spins so slowly and how magnetism affects the Sun's

atmosphere. Then come the evolution of stars and the unravelling of the way in which stellar compositions changed with time. This is followed by the mysteries concerning the methods of galaxy and galactic arm formation and the distribution of galaxies in space. And the book ends with the perennial teaser: why did the Universe form in the first place?

What is also so 'classic' about this book is its style. Hoyle does not introduce any mathematical equations. He simply writes for the interested citizen with a clarity that is honed by the discipline of university lecturing and BBC radio broadcasting.

Throughout the text Hoyle is mindful of that unwritten convention that irritated his Cambridge colleague Ray Lytleton: apparently it was then (and still is) bad form for an observational astronomer to be challenged on grounds of competence, whereas theoreticians were routinely challenged in ways that implied incompetence almost to the point of imbecility. All I can say is that, on the evidence of *Frontiers of Astronomy*, Hoyle must have been challenged less than most. This book still encompasses all the joys and excitement of the astronomer's exploration of the scientific frontiers. Astronomy should be all about struggling with the great questions. It is like a good detective story — there is some vital clue that should reveal the solution to the mystery, but the clue and its significance are far from obvious.

Hoyle is an astronomer's astronomer. To quote from the first chapter of *Frontiers of Astronomy*: "Man's claim to have progressed far beyond his fellow animals must be supported not by his search for food, warmth and shelter (however ingeniously conducted), but by his penetration into the very fabric of the Universe. It is in the world of ideas and in the relation of his brain to the Universe itself, that the superiority of Man lies. The rise of Man may justly be described as an adventure in ideas."

Read this book and prepare to join the endeavour.

David Hughes is in the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.

A geometric technique for relocating hotspots and refining absolute plate motions

Paul Wessel & Loren Kroenke

School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 2525 Correa Road, Honolulu, Hawaii 96822, USA

An age-independent, geometric relationship is presented that links hotspots to the seamounts which they produce, and so permits the use of undated seamounts to refine the motion of tectonic plates. This technique has the potential to rigorously assess hotspot fixity and to locate extinct hotspots. The present application of this method points to a recent change in Pacific plate motion, and suggests a relocation of the Louisville hotspot to the Hollister ridge, south of the Eltanin fracture zone.

Seamount chains play an important role in deciphering the past motions of the Earth's oceanic plates. Following Wilson's¹ suggestion that the Hawaiian Islands were formed by the movement of the Pacific plate over a hotspot deep in the Earth's mantle, Morgan² first noted the congruence of the major island and seamount chains in the Pacific, suggesting that hotspots are stationary or fixed with respect to the mantle. As a plate moves over a stationary hotspot, chains of seamounts form that record the history of the motion. If radiometric dates for the seamounts are available one can reconstruct this history. Ever since Morgan first formulated the hotspot hypothesis, the distribution of intraplate seamount ages and the geometry of the chains have been used to determine rigid plate motions over presumably stationary hotspots (see, for example, refs 3 and 4).

Some seamounts, however, appear to be formed in the proximity of divergent plate boundaries and adjacent fracture zone systems⁵, unrelated to hotspot activity. Because seamounts can be produced by upwelling at mid-ocean spreading ridges as well as by hotspot volcanism, it has been traditionally required that seamounts believed to be hotspot-produced must 'back-track' to a known hotspot location. This step necessitates knowledge of the history of plate motion (conveniently expressed by a set of Euler stage poles that describe how spherical caps move on a spherical Earth) as well as the age of the seamount. Unfortunately, less than 1% of intraplate seamounts have been dated, and given the high cost of acquiring new samples it is unlikely we will ever attain substantial improvement in coverage. Because both the geometry of seamount chains and radiometric dates become more difficult to decipher as one goes back in time, the full potential of hotspot-trail reconstructions to decode the absolute motions of plates remains untapped. Indeed, it has even been suggested that considerable motion may have occurred between hotspots⁶. Although several studies have placed limits on how fast hotspots may migrate^{7,8}, the scarcity of dated seamounts with reliable ages renders such results inconclusive⁹. Recently, both the assumption of hotspot fixity as well as the hotspot hypothesis itself have come under renewed scrutiny¹⁰⁻¹².

We have discovered a geometric relationship between hotspot-produced seamounts and the hotspot location that allows us to determine hotspot locations from seamount locations without requiring seamount age information. The present location of a hotspot-produced seamount and a set of stage poles determine the path that the sea floor under the seamount followed over the mantle from the ridge crest to its present location. The (stationary) hotspot must be located somewhere along this flow line. All seamounts produced by a particular hotspot will have flow lines that intersect at the unique hotspot location. We use this principle to refine recent

Pacific plate stage poles and relocate the Louisville hotspot. The new location is close to the source region for a recent swarm of intense T waves (indicative of eruptions or shallow earthquakes) and a local geoid high; recent dredging at this site recovered zero-age rocks. Finally, the proposed new direction of absolute motion may explain the oblique trend of the Marquesas Islands, and the spreading centre propagation and plate reorganizations at the East Pacific Rise and the Pacific–Antarctic ridge.

Decoupling geometry from age progression

Conventional 'back-tracking' of seamounts requires knowledge of the seamount's age and the stage poles that describe the plate's motion in the hotspot reference frame. We now consider Suiko seamount in the Emperor chain (Fig. 1). Given its age of 62.8 Myr, we use the relevant stage poles to 'back-track' the seamount to its hotspot origin (grey star); the path that results (heavy dashed line) is of geological interest because it delineates the expected area of volcanism. In the absence of radiometric dates, a seamount's age can

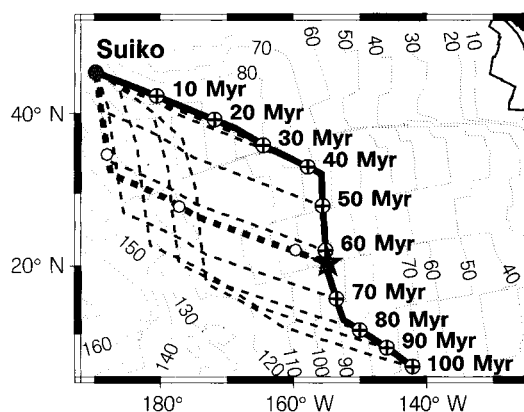


Figure 1 Calculated theoretical hotspot locations (crossed circles) for Suiko (grey circle), a dated (62.8-Myr) seamount in the Emperor chain. These locations were calculated on the basis that seamounts must be older than 0 Myr and younger than the age of the crust they sit on. Contours show crustal ages in 10-Myr intervals³⁹. The locus of theoretical hotspot locations (heavy solid line) corresponds to the trajectory the sea floor beneath Suiko followed over the mantle, that is, the crustal flow line. If the hotspot (grey star) is stationary it must lie somewhere along this path. The actual seamount trail is shown by heavy dashed line; open circles are three other seamounts in the same chain. This path has a more intuitive geological appeal (the locus of hotspot volcanism) than the flow line.

only be bracketed between 0 and the age of the crust beneath the seamount. Thus, if Suiko had not been dated, its age could be anywhere between 0 and 100 Myr. By systematically assigning possible ages from 0 to 100 Myr to Suiko we can use the stage poles to determine where the theoretical hotspot location would have been. We do so by using the conventional 'back-tracking' described above for these hypothetical ages; each 'back-track' path (dashed lines) terminates at the corresponding hypothetical hotspot location. The locus of these positions defines all possible locations of a stationary Hawaiian hotspot (heavy black line), and includes the actual hotspot location (grey star). On the other hand, the traditional 'back-track' path (heavy dashed line) only includes the hotspot when the seamount age is known.

It occurred to us that the solid line is a sea-floor flow line; that is, the trajectory of crust formed at a particular point on a mid-ocean ridge (whose palaeo-location coincides with the 100 Myr symbol in Fig. 1). Therefore, there are no features on the sea floor along most of this line (other than by pure coincidence). Consequently, traditional 'back-tracking' does not retrace the actual trajectories of a seamount (solid line segment from hotspot to seamount) but instead follows it back along the seamount chain (heavy dashed line). In other words, a path starting from a hotspot and directed along a seamount chain to a certain seamount does not give the

displacement history of the seamount relative to the hotspot, except in the case where no change in the direction of plate motion has occurred. Given Suiko's age of 62.8 Myr, both the heavy solid and dashed lines lead to the hotspot. Mathematically, the two paths reflect the same set of stage rotations applied in the opposite order; their difference is thus a consequence of the non-commutivity of finite rotations.

The usefulness of flow lines becomes clearer when more than one seamount is considered (Fig. 2). Given the crustal age beneath four Hawaiian–Emperor seamounts, we construct each seamount location's flow line (Fig. 2 bottom, red lines). Because one point along each line coincides with the hotspot location, it follows that the four paths must intersect at that location. On the other hand, Johnston Atoll and Brahms seamounts, both unrelated to the Hawaiian hotspot, have individual flow lines (Fig. 2 bottom, green lines) that do not intersect the red lines at the hotspot. The terminal points represent the location of the ridge crest (with respect to the mantle) when the sea floor under these seamounts was formed.

Though we chose the age of the crust as an upper limit in the example above, it now becomes clear that crustal ages are merely beneficial but not necessary to the method. In the absence of reliable crustal ages, we simply assign a seamount the age of what we believe is much older sea floor (for example, 145 Myr). This simply extends

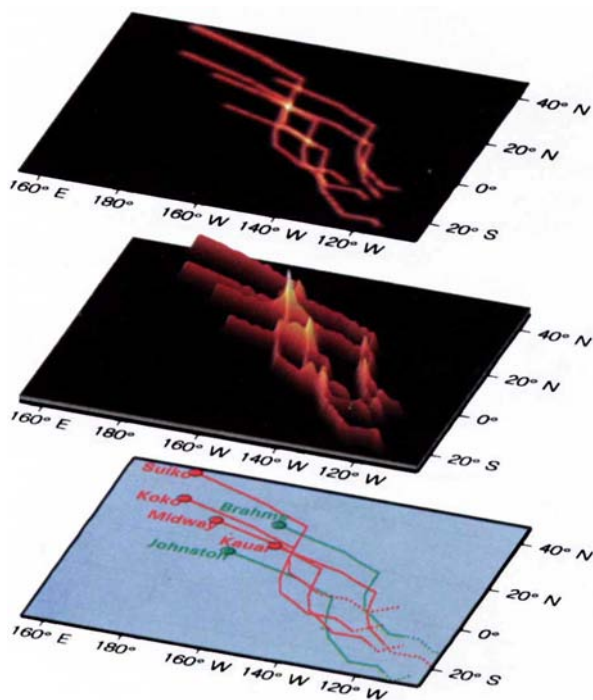


Figure 2 Bottom panel, flow lines (red) of Suiko, Koko, Midway and Kauai seamounts, which all belong to the Hawaii–Emperor chain. These lines were drawn by using the age of the crust beneath the seamounts as an upper limit. Because each flow line contains the true hotspot location, it follows they must all intersect at the hotspot. Other seamounts (for example, Johnston Atoll and Brahms seamount) have flow lines (green) that generally go elsewhere. The terminal points on these paths represent the location of the palaeoridge. The upper age limit is arbitrary: setting it to 145 Myr simply extends the mantle trajectories slightly further back in time (dashed lines) and does not change the intersection at the hotspot. Middle panel, the geometry based on line intersections extended to allow for a finite width and height for each seamount. The width and height can be measured from bathymetry or inferred from satellite gravity data. Convolution of seamount shapes with flow lines gives a cumulative volcano amplitude (CVA) surface. Top panel, CVA surface presented as an image. Flowline convergence over hotspots will cause bright spots (local maxima) in the CVA image (as will random flow line interference).

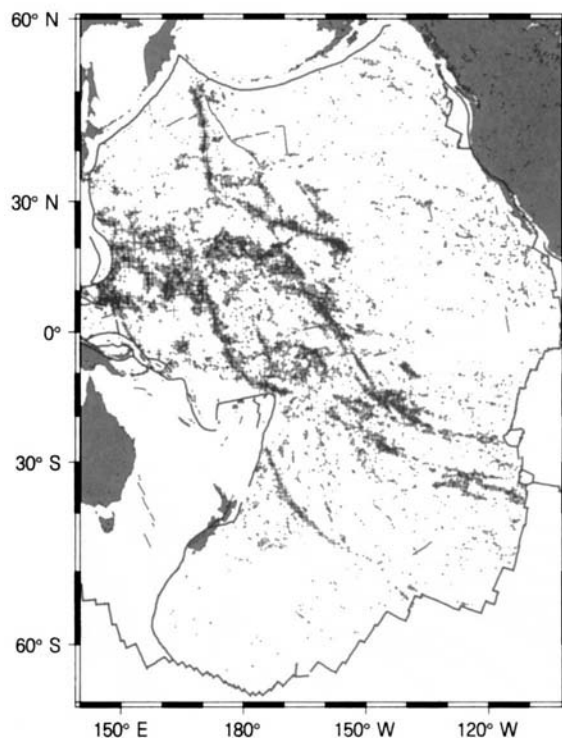


Figure 3 Locations of 8,800 seamounts determined automatically from remotely sensed gravity data. Local maxima in the vertical gravity gradient were determined and modelled as gaussian-shaped seamounts of fixed aspect-ratio; by tracing the zero contour and finding all local maxima inside it, we were able to assign a radius and amplitude to each seamount. Certain limitations in the automated process determine how small a seamount we can detect; currently, only seamounts taller than 1.5–2 km have been characterized. Features within 25 km of fracture zones (dotted lines) or troughs were excluded. The size of the symbols is proportional to the maximum in the vertical gravity gradient field. This preliminary data set, however, still contains some volcanic ridges and tectonic features mischaracterized as seamounts.

the flow lines back in time (Fig. 2 bottom, dashed red and green lines); it does not affect the intersection of flow lines over a hotspot. Consequently, our decoupling of geometry and age progression means that instead of being limited to a few hundred seamounts with dates of highly variable quality, we may employ more than 10,000 seamounts that can be characterized remotely using satellite altimetry¹³.

The final step is to extend the point-and-line geometry to seamounts of finite radius and height. Instead of finding the intersections of lines of negligible thickness we convolve each seamount's bathymetric or gravimetric expression with its flow line to obtain an elongated ridge of finite width and height. Figure 2 middle shows the linear ridges resulting from the convolution and how they combine to produce a local maximum over the hotspot. The sum of these ridges results in a surface of cumulative volcano amplitude (CVA). At hotspots the CVA value reflects the cumulative volume produced by that hotspot, allowing us to assess the 'strength' of each hotspot. The CVA also allows larger seamounts to contribute relatively more to the determination of the hotspot locations. The top panel in Fig. 2 shows CVA values presented as an image. Maxima in the CVA image correspond to the line intersections below. We refer to the technique of finding hotspots via maxima in the CVA image as 'hot-spotting'.

Relocating Pacific hotspots

We have used the satellite-derived gravity field¹⁴ to automatically locate and characterize Pacific intraplate seamounts remotely, resulting in a data base of 8,800 seamounts¹⁵ (Fig. 3). Figure 4a shows the CVA image for Hawaii using this database and a recent, modified set of stage poles^{16–18}. The CVA maximum occurs at a distinct intersection north of the big Island of Hawaii, 135–180 km away from the sites of current volcanic activity. Because of the distinct bend in the Hawaiian Island chain near Molokai, several authors have proposed a recent change in absolute plate motion^{19–22}. Moreover, because the location of the Hawaiian hotspot is assumed to be well determined from seismicity and active volcanism at Loihi and Kilauea, we attempted to find a stage pole for the last few million years that would explain, to first order, the apparent recent bend in the trend of the Hawaiian Islands, and that would relocate the maxima in the CVA image to coincide with the 'known' location of the Hawaiian hotspot (that is, near Loihi and Kilauea). Although changes could have occurred between 1 and 5 Myr, by trial-and-error we settled on a change in plate motion at ~3 Myr, giving us a stage pole (Table 1) which satisfied our requirements (Fig. 4b). Our new stage pole has important implications for hotspots and hotspot-trails elsewhere in the Pacific. The hot-spotting technique discussed above relocates the Louisville hotspot ~375 km further

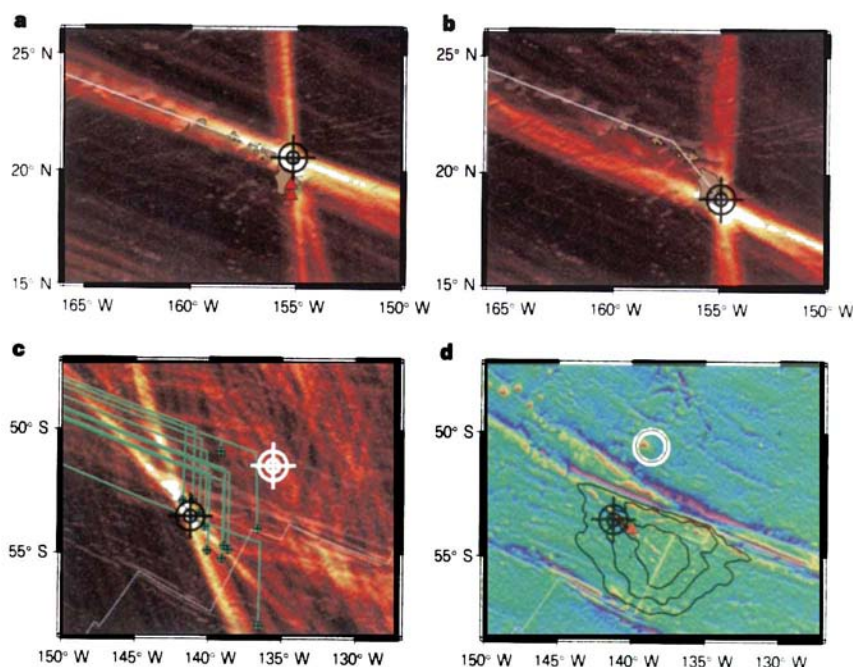


Figure 4 a, CVA image for the Hawaiian hotspot with artificial illumination using the slopes in the gravity field. Without a recent change in plate motion the CVA maximum (crossed circles), and hence the best hotspot location, is far removed from the centre of current volcanic activity at Loihi/Kilauea (red triangles). **b**, CVA image after allowing for a change in absolute plate motion near Molokai. The CVA maximum is now <35 km from Loihi/Kilauea. **c**, CVA image for the Louisville region locates the hotspot (black crossed circles) at 141°12' W, 53°33' S; the white crossed circles indicate the location without the recent plate motion change. The green circles represent the 'back-tracked' locations of dated seamounts from the Louisville chain; they generally agree with our location but show much scatter due

to age uncertainties (both analytical and geological). **d**, Gravity field over the Louisville hotspot area. Our hotspot location (crossed circles) is <75 km from a site of underwater eruptions (red triangle) located on the Hollister ridge which extends from the mid-ocean ridge to the Eltanin fracture zone. The ridge rises to within ~100 m of the sea surface. A local geoid high (solid lines are 0.5-m contours) is centred on this ridge segment and is elongated in the direction of the Hollister ridge. The large seamount inside the circle has a radiometric age of 0.5 Myr and was believed to be close to the hotspot²⁴, but in light of our reconstruction it probably was never directly over the hotspot.

Table 1 Pacific plate absolute motion stage poles

Long. (deg.)	Lat. (deg.)	Interval (Myr)	Opening (deg.)
218.63	− 3.44	145–125	− 9.8
145.17	− 18.08	125–110	− 8.3
184.40	− 50.80	110–100	− 7.4
319.50	18.90	100–74	13.4
265.00	22.00	74–65	7.5
248.59	9.67	65–54.2	6.1
261.79	16.19	54.2–43.1	6.8
297.52	57.17	43.1–19.9	14.4
353.10	67.26	19.9–15.1	4.0
272.50	71.00	15.1–3.0	13.2
333.00	25.00	3.0–0	3.6

south (Fig. 4c) than commonly suggested^{23,24}. Conventional ‘back-tracking’ of the few dated seamounts in the Louisville chain is in general agreement with our geometric determination. Interestingly, our location is only 75 km from the source region for intense, shallow seismicity or eruptions (as suggested by a swarm of strong, monochromatic T-wave activity) on the Hollister ridge²⁵, over 150 km away from the Pacific–Antarctic ridge. The Hollister ridge is 450 km long, 20 km wide, and rises to within ~100 metres of sea level; zero-aged rocks were dredged at this site²⁶. Moreover, the Hollister ridge and the segment of the mid-ocean ridge it abuts are within a region of elevated geoid anomalies characteristic of hotspot swells (Fig. 4d). Overall, the Hollister ridge has the appearance of being a result of hotspot–ridge interaction²⁷, similar to Necker ridge near the Hawaiian chain and the Wolfe–Darwin lineament near the Galápagos Islands.

The suggested change in plate motion implies that seamount chains produced by the multiple hot spots in French Polynesia should show evidence of major over-printing and confusing age-progressions. Indeed, this appears to be the case^{28,29}. The recent change in plate motion is also consistent with the oblique trend of the Marquesas Islands. Future studies will attempt to clarify the relationships between the seamount chains and the hotspots that may have produced them. We also note that there is ample evidence for massive reorganization and propagation at the East Pacific Rise³⁰ and the Pacific–Antarctic ridge^{31,32} in the Late Neogene, between the onset of rifting in the Lau basin³³ and Baja California³⁴ (5–6 Myr ago) and the reorientation of Lau basin spreading just before 2 Myr ago³⁵. This activity is an expected consequence of one or more changes in plate motion in the 1–5 Myr interval, perhaps partly in response to the collisions of the Ontong Java plateau with the Solomon arc³⁶, and the New Hebrides arc with the Tonga arc³⁷.

Implications for the hotspot hypothesis

Applying the hot-spotting method to all seamounts in Fig. 3 using the refined Pacific plate stage poles (Table 1) we obtained the CVA image shown in Fig. 5. A very clear ‘X’ marks the location of the Hawaiian hotspot and illustrates the power of our new technique. A prominent oblique crossing also indicates the location of the Louisville hotspot. Other conventional hotspots in areas where numerous flow lines from the West and Central Pacific converge are not revealed by such obvious signatures. As Fig. 5 shows, only the Hawaii and Louisville hotspots are well resolved in the CVA image. Indeed, taken at face value, the CVA image may lead one to question the validity of the hotspot hypothesis. Both Rarotonga and Rurutu are located near but not on a large, broad CVA intersection. Furthermore, the locations of other proposed hotspots such as Macdonald, Society and Pitcairn do not coincide with significant CVA maxima. There are at least seven possible explanations for this: (1) The published locations for hotspots in this region may not be

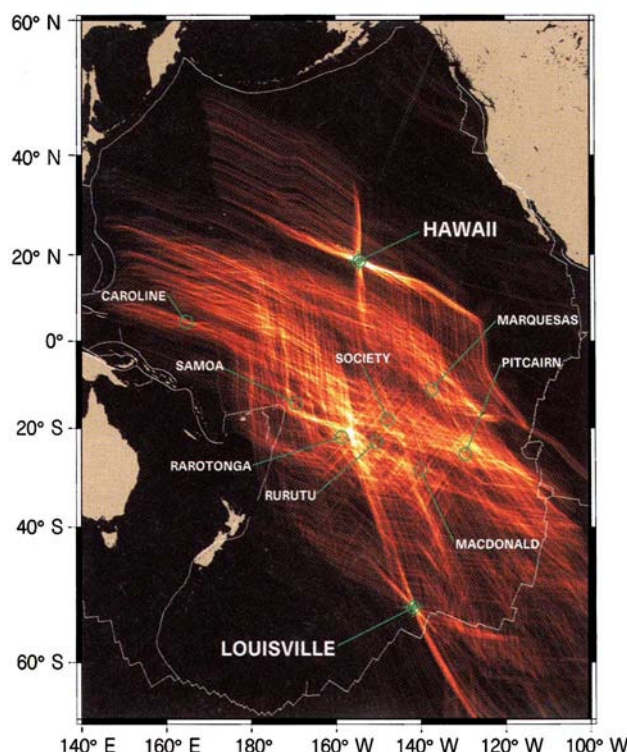


Figure 5 CVA image for the entire Pacific plate, with prominent intersections at Hawaii and Louisville. Other hotspot locations suggested in the literature are also

indicated but, for the reasons discussed in the text, may not correlate well with local CVA maxima.

accurate (as in the Louisville example); (2) the blurred CVA image may mean that our stage poles are not accurate enough; (3) it is possible that the blurred image reflects actual migration of some hotspots with respect to the Hawaii and Louisville hotspots (which show little or no relative motion, as was also recently documented by other methods¹⁸); (4) intraplate deformation may have taken place, meaning the several platelets comprising the Pacific plate have not moved as one rigid segment; (5) our seamount database is preliminary and contains numerous 'seamounts' likely to be part of volcanic ridges or tectonic features; (6) a series of contemporaneous hotspots aligned with a flow line will produce massive over-printing in the CVA image; and (7) flow lines for seamounts on old crust cause interference with other flow lines in regions away from hotspots. Such interference can be diminished by removing seamounts attributed to a known hotspot; this will also reduce the over-printing problem. Careful manual characterization of the seamounts visible in the satellite-derived gravity, aided by ship-board data where available, will result in an optimal database for future investigations. However, possibilities (2) and (3) are the most exciting: if the hotspots are stationary then we can refine the stage poles until the image becomes focused. Simultaneously, we may solve for the amount of motion (if any) of each hotspot that will likewise sharpen the image. Thus, for the first time we have a method with a potential to rigorously address the issue of hotspot fixity using a large data set.

Because Fig. 5 is a product of flow lines and gravimetric signatures, any short-lived or recently formed hotspots would be difficult to detect. Although seamounts produced before and after a pronounced change in plate motion give rise to spectacular intersections in the CVA image, our technique should also work even for segments without significant bends, provided the seamount chain extends all the way to the location of the (present or extinct) hotspot. We note that the short Caroline hotspot trail gives a small, oblique CVA intersection whose maxima coincides closely with the proposed location of the Caroline hotspot³⁸, and the Marquesas hotspot is conspicuous at the northwestern end of an elongated but poorly resolved CVA maximum.

Being independent of age information, the hot-spotting technique can only resolve the geometry of plate motion (that is, stage pole opening angles). But by incorporating available seamount age information we will be able to assign opening rates to each stage pole and use this model to predict seamount ages. In making the CVA images, we used predicted seamount radius and gravimetric amplitude¹⁵. Alternatively, by using predicted seamount total volume (including the compensating root) as amplitude and assuming the predicted seamount ages are valid we suggest it may be possible to map temporal and spatial variations in hotspot volcanism since the Early Mesozoic.

Although the above lines of evidence provide intriguing support for both a recent change in plate motion (~3 Myr ago) and a more southerly location of the Louisville hotspot, we note these results are preliminary and are likely to change when a better reconstruction method is used, that is, one that simultaneously determines the best stage poles as well as the motion, if any, of the hotspots. Nevertheless, with an increase of almost two orders of magnitude in the number of usable data (seamount locations) it may be possible to attain significant advances in four areas: the quality of plate reconstructions, the rigorous reassessment of the fixity of hotspots, the discovery of locations of extinct hotspots (for example, the

Musician hotspot), and the deciphering of temporal variations in intraplate volcanism. □

Received 15 January; accepted 8 April 1997.

- Wilson, J. T. A possible origin of the Hawaiian islands. *Can. J. Phys.* **41**, 863–870 (1963).
- Morgan, W. J. Convection plumes in the lower mantle. *Nature* **230**, 43–44 (1971).
- Jarrard, R. D. & Clague, D. A. Implications of Pacific island and seamount ages for the origin of volcanic chains. *Rev. Geophys.* **15**, 57–76 (1977).
- Duncan, R. A. & Clague, D. A. in *The Ocean Basins and Margins* (eds Nairn, A. E. M., Stehli, F. G. & Uyeda, S.) 89–121 (Plenum, New York, 1985).
- Batiza, R. & Vanko, D. Volcanic development of small oceanic central volcanoes on the flanks of the East Pacific Rise inferred from narrow-beam echo-sounder surveys. *Mar. Geol.* **54**, 53–90 (1983).
- Molnar, P. & Stock, J. Relative motions of hotspots in the Pacific, Atlantic and Indian Oceans since Late Cretaceous time. *Nature* **327**, 587–591 (1987).
- Acton, G. D. & Gordon, R. G. Paleomagnetic tests of Pacific plate reconstructions and implications for motion between hotspots. *Science* **263**, 1246–1254 (1994).
- Müller, R. D., Royer, J.-Y. & Lawyer, L. A. Evidence for hotspot motion in the Late Cretaceous/Early Tertiary. *Geol. Soc. Am. Abstr. Progn.* **23**, A318 (1991).
- Duncan, R. A. & Richards, M. A. Hotspots, mantle plumes, flood basalts, and true polar wander. *Rev. Geophys.* **29**, 31–50 (1991).
- Anderson, D. L. The perisphère-PLUME model. *Eos* **77**, F769 (1996).
- Norton, I. O. Plate motions in the North Pacific: The 43 Ma nonevent. *Tectonics* **14**, 1080–1094 (1995).
- Tarduno, J. A. & Gee, J. Large-scale motion between Pacific and Atlantic hotspots. *Nature* **378**, 477–480 (1995).
- Craig, C. H. & Sandwell, D. T. Global distribution of seamounts from Seasat profiles. *J. Geophys. Res.* **93**, 10408–10420 (1988).
- Smith, W. H. F. & Sandwell, D. T. Marine gravity field from declassified Geosat and ERS-1 altimetry. *Eos* **76**, F156 (1995).
- Lyons, S. & Wessel, P. Determination of seamount distribution on the Pacific plate from satellite altimetry. *Eos* **77**, F315 (1996).
- Kroenke, L. W. & Sager, W. W. The formation of oceanic plateaus on the Pacific plate. *Eos* **74**, 555 (1993).
- Yan, C. Y. & Kroenke, L. W. A plate tectonic reconstruction of the southwest Pacific, 0–100 Ma. *Proc. ODP Sci. Res.* **130**, 697–709 (1993).
- Yan, C. Y. *Evaluating the fixity of Hawaiian and Louisville hotspots and related seamounts and the rigidity of the Pacific plate*. Thesis, Texas A&M Univ. (1996).
- Harbert, W. & Cox, A. Late Neogene motion of the Pacific plate. *J. Geophys. Res.* **94**, 3052–3064 (1989).
- Pollitz, F. Pliocene change in Pacific-plate motion. *Nature* **320**, 738–741 (1986).
- Cox, A. & Engerbreton, D. Change in motion of the Pacific plate at 5 Ma. *Nature* **313**, 472–474 (1985).
- Epp, D. Possible perturbations to hotspot traces and implications for the origin and structure of the Line Islands. *J. Geophys. Res.* **89**, 11273–11286 (1984).
- Watts, A. B., Weissel, J. K., Duncan, R. A. & Larson, R. L. Origin of the Louisville ridge and its relationship to the Eltanin fracture zone system. *J. Geophys. Res.* **93**, 3051–3077 (1988).
- Lonsdale, P. Geography and history of the Louisville hotspot chain in the Southwest Pacific. *J. Geophys. Res.* **93**, 3078–3104 (1988).
- Talandier, J. & Okal, E. A. Monochromatic T waves from underwater volcanoes in the Pacific ocean: Ringing witnesses to geyser processes? *Bull. Seismol. Soc. Am.* **86**, 1529–1544 (1996).
- Geli, L., Bougault, H. & the Pacarctic Scientific Party *La campagne Pacarctic du N/O L'Atlantique* (Ifremer, Brest, France, 1996).
- Small, C. Observations of ridge-hotspot interactions in the southern ocean. *J. Geophys. Res.* **100**, 17931–17946 (1995).
- McNutt, M. K., Caress, D. & Reynolds, J. Multiple episodes of lineated midplate volcanism in the southern Austral islands. *Eos* **77**, F769 (1996).
- Turner, D. L. & Jarrard, R. D. K-Ar dating of the Cook-Austral island chain: A test of the hot-spot hypothesis. *J. Volcanol. Geotherm. Res.* **12**, 187–220 (1982).
- Bird, R. T. & Naar, D. F. Intraplate origins of mid-ocean ridge microplates. *Geology* **22**, 987–990 (1994).
- Cande, S. C., Raymond, C. A., Stock, J. & Haxby, W. F. Geophysics of the Pitman fracture zone and Pacific-Antarctic plate motions during the Cenozoic. *Science* **270**, 947–953 (1995).
- Cande, S. C. & Kent, D. V. A new geomagnetic polarity time scale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* **97**, 13917–13951 (1992).
- Hawkins, J. W. Petrologic synthesis: Lau basin transect (Leg 135). *Proc. ODP Init. Rep.* **135**, 879–905 (1994).
- Larson, R. Bathymetric, magnetic anomalies, and plate tectonic history of the mouth of the Gulf of California. *Geol. Soc. Am. Bull.* **83**, 3345–3360 (1972).
- Taylor, B., Zellmer, K., Martinez, F. & Goodliffe, A. Sea-floor spreading in the Lau back-arc basin. *Earth Planet. Sci. Lett.* **144**, 35–40 (1996).
- Cande, S. C., Haxby, W. F. & Kent, D. V. Synchronous changes in plate motion in the Late Neogene. *Eos* **72**, 99 (1991).
- Kroenke, L. W. in *The Origin and Evolution of Pacific Island Biota* (eds Keast, A. & Miller, S. E.) 19–34 (SPB Academic, Amsterdam, 1996).
- Keating, B. H. et al. Evidence for a hot-spot origin of the Caroline islands. *J. Geophys. Res.* **89**, 9937–9948 (1984).
- Müller, R. D., Roest, W. R., Royer, J. Y., Gahagan, L. M. & Sclater, J. G. Digital isochrons of the world's ocean floor. *J. Geophys. Res.* **102**, 3211–3214 (1997).

Acknowledgements. We thank G. Acton, D. Bercovici, J. Mahoney, S. Martel, M. McNutt and S. Stein for their reviews. This work was supported by the US National Science Foundation.

Correspondence should be addressed to P.W. (e-mail: wessel@soest.hawaii.edu).

Structure of ADP·AlF₄⁻-stabilized nitrogenase complex and its implications for signal transduction

Hermann Schindelin*, Caroline Kisker*, Jamie L. Schlessman*, James B. Howard† & Douglas C. Rees*

* Division of Chemistry and Chemical Engineering, 147-75CH, California Institute of Technology, Pasadena, California 91125, USA

† Department of Biochemistry, University of Minnesota, Minneapolis, Minnesota 55455, USA

The coupling of ATP hydrolysis to electron transfer by the enzyme nitrogenase during biological nitrogen fixation is an important example of a nucleotide-dependent transduction mechanism. The crystal structure has been determined for the complex between the Fe-protein and MoFe-protein components of nitrogenase stabilized by ADP·AlF₄⁻, previously used as a nucleoside triphosphate analogue in nucleotide-switch proteins. The structure reveals that the dimeric Fe-protein has undergone substantial conformational changes. The β-phosphate and AlF₄⁻ groups are stabilized through intersubunit contacts that are critical for catalysis and the redox centre is repositioned to facilitate electron transfer. Interactions in the nitrogenase complex have broad implications for signal and energy transduction mechanisms in multiprotein complexes.

Biological nitrogen fixation is catalysed by the enzyme nitrogenase, which consists of two component proteins, the Fe-protein and the MoFe-protein (reviewed in refs 1–6). As part of the overall reaction, the two proteins associate in a transient complex in which transfer of each electron is coupled to the hydrolysis of two ATP molecules. The role of nucleotide hydrolysis in this process has been compared to other nucleotide-dependent transduction mechanisms, yet nitrogenase is unique in that both proteins are required for ATP hydrolysis. Consequently, a detailed structural analysis of this complex is of considerable interest, not only for understanding the nitrogenase reaction, but also for implications as to how signal transduction occurs in multiprotein complexes. The two-component protein complex of nitrogenase is inhibited and stabilized by aluminium fluoride (AlF₄⁻) in combination with ADP^{7,8}. AlF₄⁻·ADP or AlF₄⁻·GDP have been previously used as analogues of the nucleoside triphosphate in nucleotide-switch proteins, in which structural evidence from myosin⁹ and heterotrimeric G proteins^{10,11} has indicated that the AlF₄⁻–nucleoside diphosphate might be mimicking the transition state during hydrolysis. Significantly, in nitrogenase AlF₄⁻·ADP is bound only by the two-protein complex; the individual proteins do not form adducts with AlF₄⁻·ADP. Likewise, Ras does not bind AlF₄⁻·GDP except as a complex with a second component, the GTPase-activating protein (GAP)¹². In contrast, other signal and energy-transduction proteins such as transducin and myosin tightly bind AlF₄⁻–nucleoside diphosphate as single components. The variations in conditions for binding these nucleotide analogues imply that there could be

differences in the way nucleotide binding and hydrolysis are coupled to the transduction process.

To investigate these differences, we have determined the structure of the ~360K (relative molecular mass, 360,000) multiprotein nitrogenase complex stabilized with ADP·AlF₄⁻. The crystal structure of this complex reveals that the Fe-protein has undergone substantial conformational changes. As a result, the β-phosphate and AlF₄⁻ groups are stabilized through interactions at the interface between the Fe-protein subunits, and the metal centres in the two proteins are positioned to facilitate electron transfer. The similarities and differences between the nitrogenase complex and other nucleotide-dependent switch proteins have broad implications for signal and energy transduction mechanisms in multiprotein complexes.

Overall structure

The structures of the individual components, including the associated metal centres, have been determined by X-ray crystallography^{13–17} and their properties (reviewed in refs 1–6) provide a useful framework for describing the nitrogenase complex. The MoFe-protein is a 232K protein with an α₂β₂ subunit composition that contains two unusual metal centres, the FeMo-cofactor and P-cluster. The FeMo-cofactor probably represents the active site for substrate reduction, whereas the P-cluster appears to serve as the intermediate electron transfer centre between the Fe-protein donor and the FeMo-cofactor (see below). The FeMo-cofactor is buried within the α-subunit, and the P-cluster is positioned at the interface

Table 1 Interactions between component proteins Av2 (Fe-protein) and Av1 (MoFe-protein)

Residues in Av2 (monomer 1)	Residues in Av1	Residues in Av2 (monomer 2)	Residues in Av1
Gly γ65 O	Lys α51 Nζ	Asp γ69 Oδ2	Lys β400 Nζ
Cys γ97 N	Ala β123 O	Cys γ97 N	Ile α123 O
Cys γ97 N	Val β124 O	Cys γ97 N	Val α124 O
Arg γ100 Nη1	Gly α157 O	Arg γ100 Nη2	Glu α120 Oε2
Arg γ100 Nη1	Glu α184 Oε1	Thr γ104 Oγ1	Glu α120 Oε2
Arg γ100 Nη2	Glu β120 Oε2	Gly γ133 N	Ile β158 O
Arg γ100 Nη2	Glu α184 Oε1	Arg γ140 Nη2	Asp β161 O
Thr γ104 Oγ1	Glu β120 Oε1	Glu γ141 Oε1	Asn β163 Nδ2
Thr γ104 Oγ1	Glu β120 Oε2	Asn γ173 O	Lys β171 Nζ
Gly γ133 N	Ile α159 O		
Glu γ141 Oε1	Lys α168 Nζ		

Hydrogen-bonded and ionic interactions are shown. Only interactions with a distance shorter than 3.3 Å have been included.

between the homologous α - and β -subunits. The Fe-protein is a 63K γ_2 homodimer containing one 4Fe:4S-cubane cluster located at the dimer interface. Conformational changes in the Fe-protein structure are evident upon binding of either Mg-ATP or Mg-ADP, but ATP hydrolysis only occurs in the transient complex formed between the two component proteins.

The structure of the two-protein nitrogenase complex from *Azotobacter vinelandii* was solved by a combination of molecular replacement and 2-fold non-crystallographic symmetry averaging procedures and refined to an R -factor of 0.208 (R_{free} of 0.238) at 3 Å resolution (see Methods). The nitrogenase complex has a subunit composition of $(\alpha\beta\gamma_2)_2$, with one Fe-protein dimer per $\alpha\beta$ -subunit pair of the MoFe-protein, so two Fe-protein dimers are bound per MoFe-protein heterotetramer. As a shorthand notation, the MoFe-protein and Fe-protein from *A. vinelandii* will be referred to as Av1 and Av2, respectively. When necessary, residues in Av1 are prefixed with α or β , as appropriate, and residues in Av2 are prefixed with γ . α -Helices in Av2 are designated as in ref. 15. The two $\alpha\beta\gamma_2$ halves of the complex are related by a non-crystallographic two-fold axis (Fig. 1a), whereas each Av2 dimer has a molecular two-fold axis

that coincides with the pseudo-two-fold axis relating the α - and β -subunits of Av1. The nitrogenase complex is elongated and has overall dimensions of 190 Å and 65 Å perpendicular to the non-crystallographic two-fold axis and 75 Å along this axis.

As predicted in proposed models of the complex based upon crosslinking experiments¹⁸ and rigid-body docking of the individual proteins^{14,19}, the 4Fe:4S cluster of Av2 and the P-cluster of Av1 are positioned along the pseudo-two-fold axis relating the two proteins within the complex. However, the surface incompatibility of Av1 and nucleotide-free Av2 necessitates large-scale conformational changes to accommodate specific interactions between the two proteins. As a result of these conformational changes, particularly in the Fe-protein structure (see below), the metal centres are much closer than predicted. The distance from the Av2 4Fe:4S cluster to the Av1 P-cluster is ~ 14 Å, measured as the closest atom-to-atom distance, which is ~ 4 Å shorter than expected from model building. Thus in the complex, the P-cluster is located essentially equidistant between the Fe-protein 4Fe:4S cluster and the FeMo-cofactor (also separated by 14 Å). This arrangement strongly suggests that electrons are transferred from the Fe-protein via the P-cluster to the

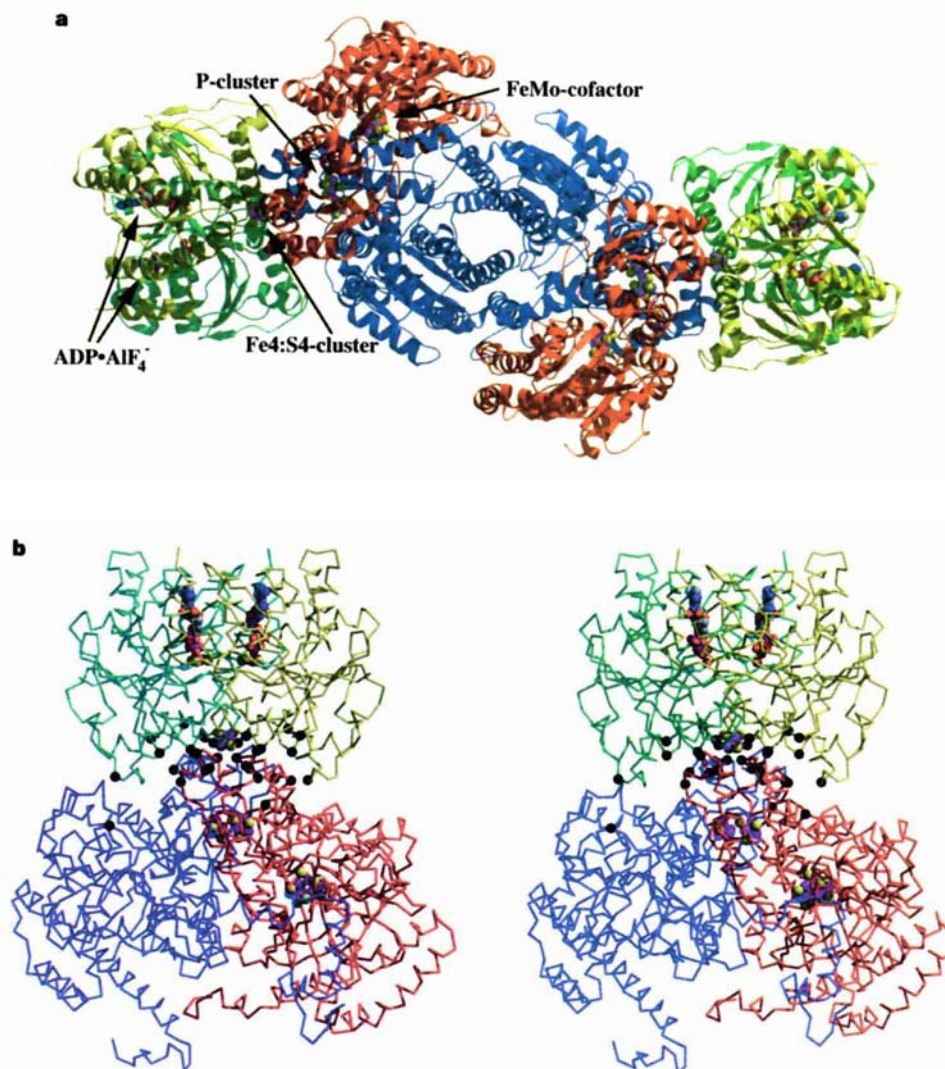


Figure 1 Overall structure of the nitrogenase complex. **a**, The entire complex with the MoFe α -subunits in red, the β -subunits in blue and the individual subunits of each Fe-protein in green and yellow. The cofactors and the bound nucleotides are shown in ball-and-stick representation. Atoms are colour-coded, with Al in orange, Mg and Mo in cyan, Fe and P in purple, S in yellow, O in red, N in blue and

C in grey. **b**, Stereo drawing of one half-complex with the subunits depicted as $C\alpha$ traces and colour-coded as in **a**. Residues participating in Av1-Av2 component interactions have their $C\alpha$ atoms highlighted as black spheres. Figures were prepared using MOLSCRIPT⁴² and RASTER3D^{43,44}.

FeMo-cofactor. In contrast to Av2, no large structural changes are detected for Av1, as reflected in an r.m.s.-deviation of 0.5 Å in Cα positions after superposition of the two MoFe-protein forms, with slightly larger r.m.s. deviations in the Av2 contact regions (see below) that comprise residues α157–α159 and β156–β158.

Interactions between component proteins

There are a large number of contacts between the two proteins in the complex (Table 1), which bury ~3,500 Å² of surface area in each Av2–Av1 interface. Most striking is the degree of intimate contact between the 4Fe:4S cluster and adjacent residues of Av2 with stretches of the polypeptide chain in Av1 (Fig. 1b). The close contact is afforded by conformational changes in Av2 that alter the environment of the 4Fe:4S cluster. The cluster is completely buried in the Av2–Av1 interface region and is in van der Waals contact with the loop regions α157–α159 and β156–β158. These interactions bring the carbonyl oxygens of Val α158 and Leu β157 to within van der Waals contact (~3.2–3.5 Å) of the Av2 4Fe:4S cluster atoms S2 and S3, respectively. The carbonyl oxygens of the pseudosymmetry-related residues α123, α124 and β123, β124 form main-chain hydrogen bonds to the amides of both Cys γ97, which provide two of the four ligands to the 4Fe:4S cluster. Other main-chain hydrogen bonds are present between the carbonyl oxygens of Ile α159/Ile β158 to the amides of Gly γ133, the residue in each γ-subunit adjacent to the second set of Cys ligands (γ132) of the Av2 cluster.

In addition to the main-chain hydrogen bond interactions near the cluster, there are numerous side-chain interactions between the two proteins in the complex. The side chain of Arg γ100 in Av2 has been considered to be of particular importance in promoting the proper binding interactions^{19–21}. Arg γ100 is in the first turn of the α3 helices that extend symmetrically from the 4Fe:4S cluster in the Av2 dimer. The side chains of both Arg γ100 protrude into two small depressions on the Av1 surface and form ionic bonds, multiple hydrogen bonds and van der Waals contacts in an asymmetric fashion to residues in the α- and β-subunits of Av1 (Fig. 2). Arg

γ100 corresponds to the site of reversible ADP ribosylation in several photosynthetic nitrogen-fixing bacteria²⁰ and the structure of the complex indicates that covalent modification of Arg γ100 by a large substituent would interrupt nitrogenase activity by steric interference with Av1. A second set of side-chain interactions involve the residues Thr γ104 from each Av2 subunit which form hydrogen bonds with either Glu α120 or Glu β120.

Structural changes in Fe-protein

The most obvious structural changes upon complex formation occur in the Fe-protein, as evidenced by the 4.1 Å r.m.s. deviation in Cα positions between the free and complexed forms of the Av2 dimer. To a first approximation, the conformational change in Av2 may be described as a ~13° rotation of each monomer toward the subunit interface that 'closes' the quaternary structure into a more compact form in the complex (Fig. 3a). Surprisingly, the 4Fe:4S cluster is not the hinge point of this motion. Instead, the cluster is shifted by ~4 Å towards the protein surface in complexed Av2. The rotation of the subunits results in a new dimer interface involving interactions between a number of highly conserved side chains. An additional measure of the subunit movement is the separation between the α-carbons of residues γ11 in the P-loops (see below) which decrease from ~10 Å in nucleotide-free Av2 to ~4 Å in complexed Av2. These conformational changes lead to a ~2 Å reduction in the calculated radius of gyration for Av2, consistent with small-angle scattering studies²² on Av2 with and without ATP, in the absence of Av1.

Five regions differ significantly when the individual subunits from isolated and complexed Av2 are compared. Most of the changes occur in α-helices and polypeptide loops surrounding the structurally conserved central β-sheet motif (Fig. 3b). If all 274 observed Cα-atoms are used to superimpose single Av2 subunits, an r.m.s.-deviation of 1.7 Å between free and complexed monomers of Av2 is obtained, whereas the deviation is only 0.8 Å if the Cα atoms of 186 core residues are considered, excluding the most altered regions. These five regions, found in both subunits, are: (1) the

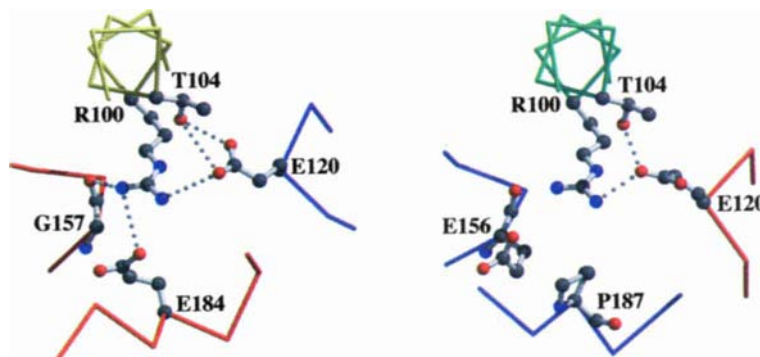


Figure 2 Interactions of Arg γ100 and Thr γ104 of Av2 with residues in Av1. The differences in contacts exhibited by these residues between the two subunits in the Av2 dimer are indicated. Subunits are designated according to the colour scheme in Fig. 1.

Table 2 Residues involved in nucleotide binding and hydrolysis

	Nitrogenase	Ras	Transducin
Sequence of Walker motif A	G-K-G-G-I-G-K-S (residues 9–16)	G-A-G-G-V-G-K-S (residues 10–17)	G-A-G-E-S-G-K-S (residues 36–43)
Sequence of Walker motif B	D-V-L-G (residues 125–128)	D-T-A-G (residues 57–60)	D-V-G-G (residues 196–199)
Residues stabilizing attacking nucleophile	D 36 (s.c.) D 129'	T 35 (m.c.) Q 61	T 177 (m.c.) Q 200
Residue stabilizing leaving group	K 10'	R from GAP	R 174

In the nitrogenase Fe-protein, residues belonging to the second subunit are indicated by a prime. GAP is the GTPase-activating protein required for fast hydrolysis of GTP by Ras. For all three proteins (the Fe-protein and the two G proteins, Ras and transducin), the residues in the Walker motifs A and B are shown with their respective sequence numbers; m.c. and s.c. indicate involvement of either main-chain or side-chain atoms in the stabilization of the transition state.

phosphate-binding loop (P-loop) or Walker A motif containing residues $\gamma 9$ to $\gamma 15$. Binding of the nucleotide results in a tightening of the turn in the P-loop. (2) Residues $\gamma 51$ to $\gamma 75$ comprising the short helix, $\alpha 2$, and adjacent residues. Structural changes in this region appear to be coupled to those in the region from $\gamma 88$ to $\gamma 118$ (see below). (3) Residues $\gamma 88$ to $\gamma 118$, including the cluster ligands Cys $\gamma 97$. The structural changes involve a large reorganization of residues $\gamma 91$ to $\gamma 97$ and the displacement of helix $\alpha 3$ away from the 4Fe:4S cluster, thereby placing the cluster more nearly in the plane

of this surface. This shift permits proper positioning of the Arg $\gamma 100$ side chains into their respective binding pockets on Av1. Contacts between the side chains of helix $\alpha 3$ and residues $\gamma 51$ to $\gamma 75$ are such that movement in one region affects the other. (4) Residues $\gamma 127$ to $\gamma 143$, including the cluster ligands Cys $\gamma 132$ and overlapping the Walker B motif Asp-X-X-Gly (residues $\gamma 125$ – $\gamma 128$). This region, which corresponds to switch II in the G-protein family³, is linked to changes involving residues $\gamma 88$ to $\gamma 118$ through formation of intersubunit main-chain hydrogen bonds near the cluster. Notable

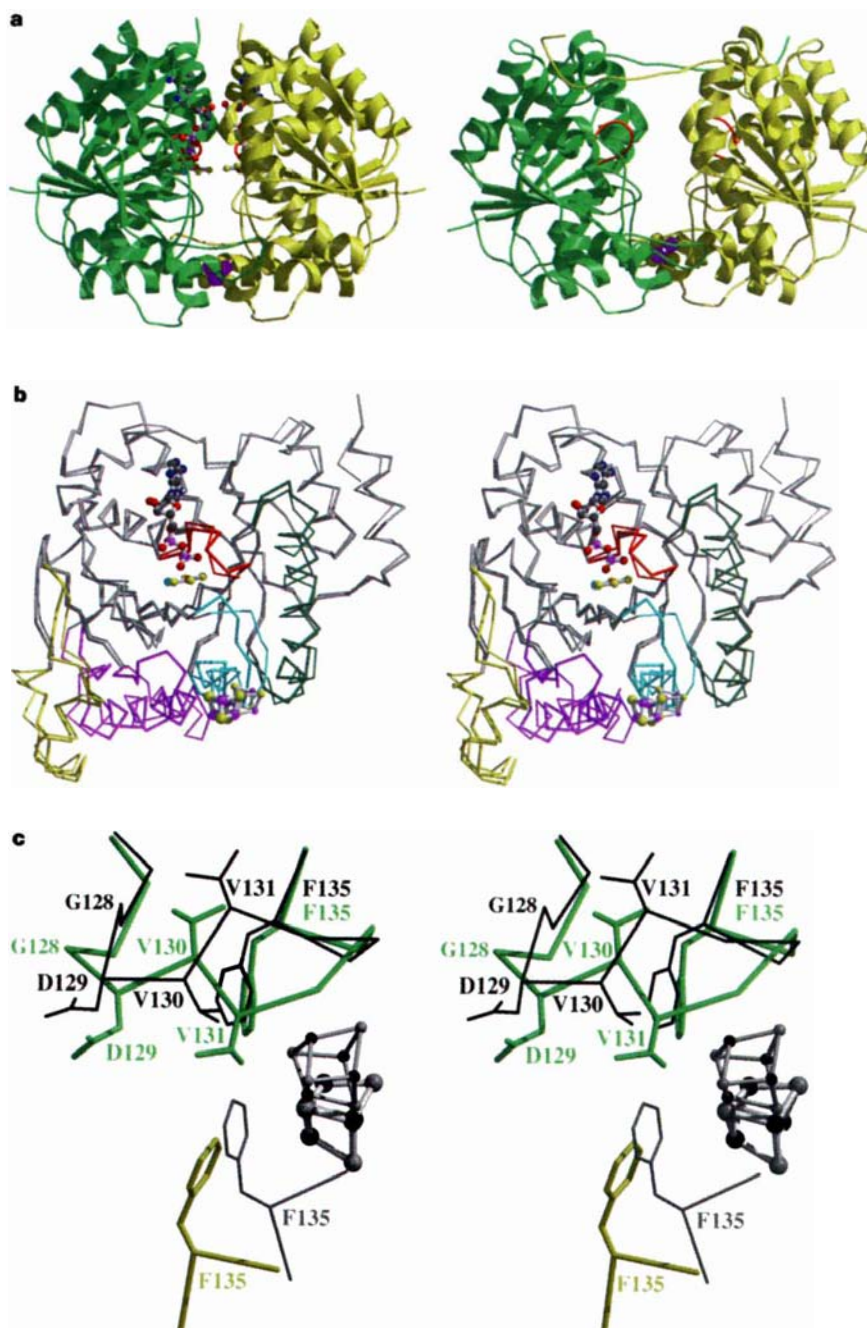


Figure 3 Structural changes in Av2. **a**, Comparison of the quaternary structures of complexed (left) and uncomplexed (right) Av2. The phosphate-binding loops, residues $\gamma 9$ to $\gamma 16$, have been highlighted in red in both structures. The C-terminal 15 residues in the complex form of Av2 are disordered and not included in the model. **b**, Stereo drawing of a least-squares superposition of uncomplexed (thin lines) and complexed (thick lines) Av2 monomers. The differing regions have been highlighted in colour: residues $\gamma 9$ – $\gamma 16$ in red, $\gamma 51$ – $\gamma 75$ in yellow, $\gamma 88$ – $\gamma 118$ in

purple, $\gamma 127$ – $\gamma 143$ in cyan and $\gamma 151$ – $\gamma 176$ in green. **c**, Detailed view near the 4Fe:4S cluster of the superposition of the isolated and complexed forms of Av2. The two monomers in complexed Av2 are shown in green and yellow thick lines; monomers in uncomplexed Av2 are black and grey. The variation in packing interactions between the conserved side chains of Val $\gamma 130$ and Val $\gamma 131$ with Phe $\gamma 135$ in the two Av2 forms are shown.

among these changes is the large movement of Gly γ 128, which forms a hydrogen bond to AlF_4^- through its amide group, and reorientation of the Val γ 130 and Val γ 131 side chains which have different packing interactions with Phe γ 135 in the free and complexed Av2 structures (Fig. 3c). (5) Residues γ 151 to γ 176 containing helix α 5. Helix α 5 is both moved and kinked at Gly γ 167. Glu γ 154, at the N-terminal end of α 5, forms an intersubunit salt bridge with Arg γ 213, stabilizing the interaction between Arg γ 213 and the bound adenosine ring (see below). Rigid body displacement of this helix is required to accommodate the nucleotide ribose.

Fe-protein and $\text{ADP}\cdot\text{AlF}_4^-$ interactions

Two $\text{ADP}\cdot\text{AlF}_4^-$ molecules are bound per Av2-dimer, with each nucleotide primarily associated with one monomer in an orientation that is roughly parallel to the dimer interface. In this orientation, the phosphate groups interact with the P-loop and the nucleoside points away from Av1, as in the G-protein family^{23,24}. However, there are other contacts to the second subunit that are very likely to be functionally significant (Fig. 4a; see below). The α,β phosphates and the AlF_4^- interact with the P-loop through backbone amide nitrogens of residues γ 14 to γ 17, as well as the side-chain oxygens of Ser γ 16 and Thr γ 17. The negative charges on the phosphates and the AlF_4^- are compensated by the P-loop Lys γ 15 and by Lys γ 41. Most important, Lys γ 10 of the second monomer interacts across the interface with the terminal oxygen of the β -phosphate. The Mg^{2+} -ion is coordinated by Ser γ 16, the β -phosphate and two of the F-ligands of the AlF_4^- . Asp γ 125, the first residue of the conserved Walker B motif Asp-X-X-Gly is hydrogen-

bonded to Ser γ 16, but is not in direct contact to the Mg^{2+} ; the peptide amide of Gly γ 128 forms a hydrogen bond to a fluoride of AlF_4^- . Deletion of Leu γ 127 in this region generates a form of Av2 that binds tightly to Av1 without nucleotide^{25,26}. Asp γ 129, adjacent to the Walker B motif, is poised to interact with the AlF_4^- of the other subunit, probably through an intermediate water molecule by analogy with the $\text{G}\alpha$ proteins. The side-chain carboxyl oxygen of Asp γ 39 corresponds structurally to the carbonyl oxygen of Thr 177 and Thr 35 in the switch I region of transducin and Ras, respectively, which each form a hydrogen bond to the attacking water molecule.

The adenosine portion is fixed by several interactions, including ribose O2' and O3' hydrogen-bonded to the Glu γ 221 side chain, and adenine N1 and N7 are hydrogen-bonded to the main-chain amide of Asp γ 214 and to the Asp γ 185 side chain, respectively. The side chain of Arg γ 213 is stacked with its guanidinium group onto the base and is held in place by an intersubunit salt bridge to Glu γ 154 (Fig. 4a). On the opposite side of the base, Val γ 217 is in van der Waals contact with the adenine. The exocyclic amino group is hydrogen-bonded to the main-chain oxygen of Pro γ 212, which may contribute to the high specificity of nitrogenase for adenine nucleotides. As expected, the side chains interacting with the bound nucleotide are highly conserved in the sequences of Fe-proteins, and substitutions of residues γ 15 (refs 27, 28), γ 16 (ref. 29), γ 125 (ref. 30) and γ 129 (ref. 25) result in significant alterations of nitrogenase activity.

Mechanistic implications

The mechanism of ATP hydrolysis by nitrogenase has substantial

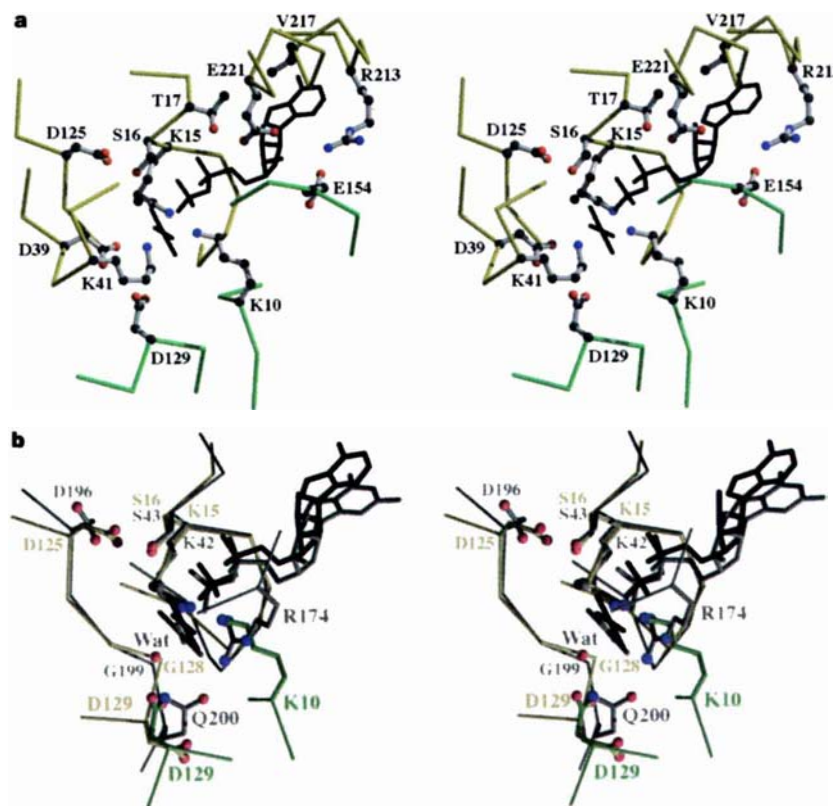


Figure 4 a, Stereo drawing of the bound nucleotide showing some of its interactions with the Fe-protein. The two subunits are depicted in yellow and green; Mg^{2+} , Al^{3+} and F^- are shown in cyan, orange and yellow, respectively; bound nucleotide is black. Side chains of residues interacting with the nucleotide are shown as ball-and-stick models and are labelled. **b**, Results of a least-squares superposition of the Walker motif A and conserved Asp-X-X-Gly sequences in Av2

and the α -subunit of transducin complexed with $\text{GDP}\cdot\text{AlF}_4^-$. Residues belonging to the individual subunits of Av2 are highlighted in yellow and green, with the bound $\text{ADP}\cdot\text{AlF}_4^-$ in black; residues belonging to transducin and bound $\text{GDP}\cdot\text{AlF}_4^-$ are grey. The bound Mg^{2+} and Ca^{2+} are shown as small black and large grey spheres, respectively, and the attacking water molecule in transducin is in red.

parallels to that deduced for G proteins³¹. The structure of the Av1–Av2–ADP·AlF₄[−] complex establishes that the binding mode of ATP to Fe-protein is very similar to that of GTP in G proteins (Fig. 4b). By analogy to Gα proteins^{10,11}, the AlF₄[−] can be considered to mimic the trigonal bipyramidal geometry of the terminal phosphate undergoing nucleophilic attack by a water molecule. Although there is some controversy (reviewed in refs 31, 32) as to whether ‘activation’ of attacking water or stabilization of the developing negative charge on the β,γ bridging oxygen atom is more important for lowering the energy of the transition state, amino-acid residues are present that might serve either function. Most notable are Lys γ10, which could stabilize the leaving group, and Asp γ129, which might activate the attacking water molecule. One unexpected feature of the nitrogenase structure, however, is that Lys γ10 and Asp γ129, needed to

complete the ensemble of residues required for nucleotide hydrolysis, are from the second subunit of Av2. Some of the amino-acid side chains in Av2 that interact with the nucleotide and their counterparts in Ras and transducin are given in Table 2. It is interesting to compare how the various nucleotide switch proteins have structurally solved the problem of providing the critical residues in a catalytically active state and perhaps of influencing the overall rate of hydrolysis. In transducin, the relevant residues Gln 200 and Arg 174 are found in different domains of the same protein^{11,24}; in Ras–GAP, Gln 61 is from Ras and a stabilizing arginine is probably contributed by a second protein, GAP³³; in Av2, Lys γ10 and Asp γ129 are from the other subunit within this dimeric protein.

The structure shows that Av1 does not directly interact with the nucleotide and hence its role in nucleotide hydrolysis must be to stabilize an Av2-nucleotide intermediate that is not attainable by Av2 alone. Because the structure presented here is only one of several potential states along the reaction coordinate, there are limitations in determining cause-and-effect relationships. However, important connections between nucleotide-related interactions and electron transfer can be made. The tertiary and quaternary structural changes in Av2 needed to complete nucleotide binding/hydrolysis likewise promote extensive interactions between the Av1 backbone and the Av2 4Fe:4S cluster. Significantly, potential electron transfer pathways can be traced between the 4Fe:4S cluster and P-cluster that would not exist without reorganization of the Av2 molecule (Fig. 5). Structural connections between the 4Fe:4S cluster and the ATP site can also be identified that involve the large movement of Gly γ128 to form a peptide backbone hydrogen bond to the AlF₄[−] and the reorientation of the side chains of Val γ130 and Val γ131. These changes occur in the loop containing Asp γ129 to Cys γ132 and directly link the nucleotide region to the cluster and the MoFe-protein interface (Figs 3c, 5). These residues correspond to the switch II region of Ras and Gα which also undergo nucleotide-mediated conformational changes^{10,11}. The main connection between ATP hydrolysis and electron transfer may be stabilization of the closed interface between the Fe-protein subunits, providing the new conformation from which these processes can proceed. Once the nucleotide is hydrolysed and phosphate has been released, the intersubunit stabilization would be decreased and the subunits could move apart, driving protein dissociation. As the nucleotide is completely buried in the ‘closed’ conformation of Fe-protein observed in the complex, formation of an ‘open’ conformation is essential for exchange of Mg-ATP for Mg-ADP.

The use of the nucleotide-switch motif to regulate electron transfer in nitrogenase demonstrates the generality and versatility of this motif, which has been recruited for various biochemical reactions. In addition, the nitrogenase complex reveals the connection between an intermediate state generated by nucleotide hydrolysis and a functional conformational state. As such, nitrogenase illustrates the ability of transient multiprotein complexes to serve as regulatory proteins by coupling nucleotide hydrolysis to signal and energy transduction processes.

Methods

Crystals of the nitrogenase complex were obtained by the anaerobic capillary batch method¹⁵ with 16–20% PEG 8000, 0.1 M cacodylate, pH 6.5, and 20–50 mM MgCl₂ as precipitant. They belong to the orthorhombic space group C222₁ (*a* = 79.0 Å, *b* = 299.7 Å, and *c* = 334.5 Å), with one α₃β₂γ₄ complex in the asymmetric unit (*M_r*, 360K). Two data sets (Table 3), Nat-1 and Nat-2, were collected from flash-frozen crystals on an R-Axis IIc imaging plate with monochromatized CuKα radiation produced by a Rigaku RU 200 rotating anode generator operating at 50 kV and 100 mA. Both data sets were processed with DENZO³⁴ and scaled with SCALEPACK. Subsequent calculations used programs from the CCP4 suite³⁵, with exceptions as indicated. The structure was solved by a combination of molecular replacement and two-fold non-

Table 3 Data collection and refinement statistics

Data collection statistics				
	Nat-1		Nat-2	
$d_{\max}$ (Å)	4.5		3.0	
Total observations	76,105		104,037	
Unique observations	26,992		57,735	
R_{sym}^*	0.153		0.098	
R_{sym} (last shell)*	0.169		0.245	
Completeness	0.813		0.738	
Completeness (last shell)	0.379		0.288	
Refinement statistics				
$R_{\text{cryst}}/R_{\text{free}}^\dagger$	0.208		0.238	
Ramachandran statistics‡	0.872	0.110	0.004	0.004
R.m.s. deviations from stereochemical target values				
r.m.s. Δ bond lengths (Å)	0.006			
r.m.s. Δ angles (°)	1.66			
r.m.s. Δ torsion angles (°)	22.12			
r.m.s. Δ improper torsion angles (°)	1.11			

**R*_{sym} = Σ_{*hkl*} Σ_{*i*} |*I_i* − ⟨*I*⟩| / Σ_{*hkl*} Σ_{*i*} *I_i*, where *I_i* is the *i*th measurement and ⟨*I*⟩ is the weighted mean of all measurements of *I*. No σ cutoff was applied.

†*R*_{cryst} = Σ_{*hkl*} ||*F_o* − *F_c*|| / Σ_{*hkl*} ||*F_o*||, where *F_o* and *F_c* are the observed and calculated structure factor amplitudes. *R*_{free} is defined as *R*_{cryst} for 2% of the data omitted in thin-resolution shells from refinement. No σ cutoff was applied.

‡ Percentage of residues in most favoured, additionally allowed, generously allowed and disallowed regions of the Ramachandran diagram according to an analysis with PROCHECK⁴⁰.

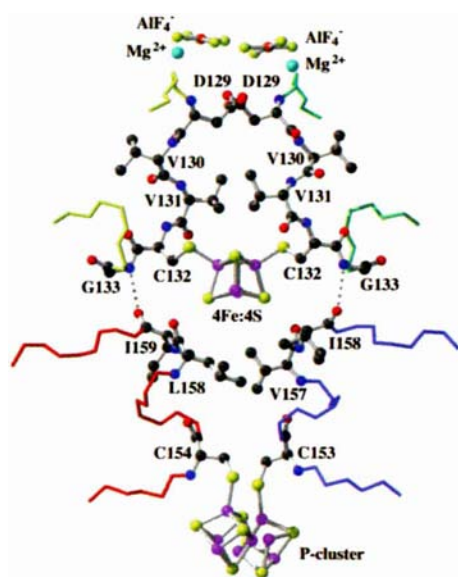


Figure 5 Signal transduction and electron transfer pathways in the nitrogenase complex. The residues shown connect the AlF₄[−] site to the P-cluster of Av1 through the 4Fe:4S cluster of Av2.

crystallographic symmetry averaging, following location of the Av1 component of the complex by molecular replacement using AMORE³⁶. Density for Av2 became interpretable after three rounds of 2-fold averaging using RAVE³⁷ at 4.5 Å resolution with data set Nat-1. The initial mask was based on the model of the complex and was updated after each round of averaging after manual rigid-body movements of first the Av2 dimer and subsequently the two Av2 monomers. At this state, data set Nat-2 became available and another round of averaging was performed to a resolution of 3.3 Å. The resulting electron density was of good quality and Av2 was completely rebuilt with the program O³⁸; however, the position of the nucleotide was not clear at this stage.

The structure was refined to 3.0 Å resolution against data set Nat-2 with the simulated annealing protocol in X-PLOR³⁹. The non-crystallographic symmetry was exploited during refinement by restraining chemically identical chains: for example, the two αβ heterodimers of Av1 and the four chains of Av2 were restrained to have almost identical conformations, resulting in pairwise r.m.s. deviations of less than 0.02 Å. Another round of averaging to 3.0 Å yielded density maps of superior quality which clearly showed the bound nucleotides. Guided by the behaviour of R_{free} , a conservative temperature factor model was included by refining one *B*-factor per residue. A bulk solvent correction allowed the inclusion of all low-resolution reflections. The current model (24,430 atoms) consists of residues 4 to 481 and 2 to 521 for the α- and β-chain of each Av1 heterodimer, residues 1 to 274 for each Av2, two FeMo-cofactors, two P-clusters, two Ca²⁺-ions, two 4Fe:4S clusters and four MgADP·AlF₄⁻ nucleotides. No attempts were made to include water molecules at this resolution. The quality of the model has been analysed with PROCHECK⁴⁰. All of the six main-chain and five side-chain parameters evaluated by the program are significantly better than usually observed at this resolution. Refinement statistics are indicated in Table 3.

Superposition of different protein structures were performed using the algorithm of Kabsch⁴¹. PDB entry 1MIN was used for the structure of the free MoFe-protein; for the nucleotide-free Fe-protein, a model was used of a new crystal form currently under refinement at 2.2 Å resolution, following molecular replacement with PDB entry 1NIP (J.L.S. unpublished).

Received 5 December 1996; accepted 26 March 1997.

- Stiefel, E. L., Coucouvanis, D. & Newton, W. E. (eds) *Molybdenum Enzymes, Cofactors and Model Systems* (Am. Chem. Soc., Washington, DC, 1993).
- Mortenson, L. E., Seefeldt, L. C., Morgan, T. V. & Bolin, J. T. The role of metal-clusters and MgATP in nitrogenase catalysis. *Adv. Enzymol. Rel. Areas Mol. Biol.* **67**, 299–274 (1993).
- Howard, J. B. & Rees, D. C. Nitrogenase: A nucleotide-dependent molecular switch. *Annu. Rev. Biochem.* **63**, 235–264 (1994).
- Peters, J. W., Fisher, K. & Dean, D. R. Nitrogenase structure and function: A biochemical-genetic perspective. *Annu. Rev. Microbiol.* **49**, 335–366 (1995).
- Burgess, B. K. & Lowe, D. J. Mechanism of molybdenum nitrogenase. *Chem. Rev.* **96**, 2983–3011 (1996).
- Howard, J. B. & Rees, D. C. Structural basis of biological nitrogen fixation. *Chem. Rev.* **96**, 2965–2982 (1996).
- Renner, K. A. & Howard, J. B. Aluminium fluoride inhibition of nitrogenase: stabilization of a nucleotide: Fe-protein: MoFe-protein complex. *Biochemistry* **35**, 5353–5358 (1996).
- Duyvis, M., Wassink, H. & Haaker, H. Formation and characterization of a transition state complex of *Azotobacter vinelandii* nitrogenase. *FEBS Lett.* **380**, 233–236 (1996).
- Fisher, A. J. et al. X-ray structures of the myosin motor domain of *Dicystostelium discoideum* complexed with MgADP·BeF₃ and MgADP·AlF₄⁻. *Biochemistry* **34**, 8960–8972 (1995).
- Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. GTPase mechanisms of G proteins from the 1.7-Å crystal structure of transducin α-GDP·AlF₄⁻. *Nature* **372**, 276–279 (1994).
- Coleman, D. E. et al. Structures of active conformers of G_{αi1} and the mechanism of GTP hydrolysis. *Science* **265**, 1405–1412 (1994).
- Mittal, R., Ahmadian, M. R., Goody, R. S. & Wittinghofer, A. Formation of a transition-state analog of the Ras GTPase reaction by Ras-GDP, tetrafluoroaluminate, and GTPase-activating proteins. *Science* **273**, 115–117 (1996).
- Kim, J. & Rees, D. C. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science* **257**, 1677–1682 (1992).
- Kim, J. & Rees, D. C. Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. *Nature* **360**, 553–560 (1992).
- Georgiadis, M. M. et al. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. *Science* **257**, 1653–1659 (1992).
- Bolin, J. T., Campobasso, N., Muchmore, S. W., Morgan, T. V. & Mortenson, L. E. in *Molybdenum Enzymes, Cofactors and Model Systems*. ACS Symposium Series No. 535 (eds Stiefel, E. L., Coucouvanis, D. & Newton, W. E.) 186–195 (Am. Chem. Soc., Washington, 1993).
- Peters, J. P. et al. Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry* **36**, 1181–1187 (1997).
- Willing, A. & Howard, J. B. Cross-linking site in *Azotobacter vinelandii* complex. *J. Biol. Chem.* **265**, 6596–6599 (1990).
- Howard, J. B. in *Molybdenum Enzymes, Cofactors and Model Systems*. ACS Symposium Series No. 535 (eds Stiefel, E. L., Coucouvanis, D. & Newton, W. E.) 271–289 (Am. Chem. Soc., Washington, 1993).
- Ludden, P. W. & Roberts, G. P. Regulation of nitrogenase activity by reversible ADP ribosylation. *Curr. Top. Cell. Regul.* **30**, 23–56 (1989).
- Wolle, D., Kim, C., Dean, D. & Howard, J. B. Ionic interactions in the nitrogenase complex. *J. Biol. Chem.* **267**, 3667–3673 (1992).
- Chen, L. et al. MgATP-induced conformational changes in the iron protein from *A. vinelandii* as studied by small-angle x-ray scattering. *J. Biol. Chem.* **269**, 3290–3294 (1994).
- Pai, E. F. et al. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution—implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**, 2351–2359 (1990).
- Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 Å crystal structure of transducin-α complexed with GTPγS. *Nature* **366**, 654–663 (1993).
- Lanzilotta, W. N., Ryle, M. J. & Seefeldt, L. C. Nucleotide hydrolysis and protein conformational changes in *Azotobacter vinelandii* nitrogenase iron protein: defining the function of aspartate 129. *Biochemistry* **34**, 10712–10723 (1995).
- Ryle, M. J. & Seefeldt, L. C. Elucidation of a MgATP signal transduction pathway in the nitrogenase iron protein: formation of a conformation resembling the MgATP-bound state by protein engineering. *Biochemistry* **35**, 4766–4775 (1996).
- Seefeldt, L. C., Morgan, T. V., Dean, D. R. & Mortenson, L. E. Mapping the site(s) of MgATP and MgADP interaction with the nitrogenase of *Azotobacter vinelandii*. *J. Biol. Chem.* **267**, 6680–6688 (1992).
- Ryle, M. J., Lanzilotta, W. N., Mortenson, L. E., Watt, G. D. & Seefeldt, L. C. Evidence for a central role of lysine 15 of *Azotobacter vinelandii* nitrogenase iron protein in nucleotide binding and protein conformational changes. *J. Biol. Chem.* **270**, 13112–13117 (1995).
- Seefeldt, L. C. & Mortenson, L. E. Increasing nitrogenase catalytic efficiency for MgATP by changing serine 16 of its Fe protein to threonine: Use of Mn²⁺ to show interaction of serine 16 with Mg²⁺. *Prot. Sci.* **2**, 93–102 (1993).
- Wolle, D., Dean, D. R. & Howard, J. B. Nucleotide-iron-sulfur cluster signal transduction in the nitrogenase iron-protein: the role of Asp¹²⁵. *Science* **258**, 992–995 (1992).
- Frech, M. et al. Role of glutamine-61 in the hydrolysis of GTP by p21^{H-ras}: An experimental and theoretical study. *Biochemistry* **33**, 3237–3244 (1994).
- Maegley, K. A., Admirals, S. J. & Herschlag, D. Ras-catalyzed hydrolysis of GTP: A new perspective from model studies. *Proc. Natl Acad. Sci. USA* **93**, 8160–8166 (1996).
- Brownbridge, G. G., Lowe, P. N., Moore, K. J. M., Skinner, R. H. & Webb, M. R. Interaction of GTPase activating proteins (GAPs) with p21^{ras} measured by a novel fluorescence anisotropy method: Essential role of Arg-903 of GAP in activation of GTP hydrolysis on p21^{ras}. *J. Biol. Chem.* **268**, 10914–10919 (1993).
- Otwiński, Z. in *Data Collection and Processing* (eds Sawyer, L., Issacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, UK, 1993).
- Bailey, S. The CCP4 suite—programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Navaza, J. AMORE—an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Kleywegt, G. J. & Jones, T. A. in *First Map to Final Model* (eds Bailey, S., Hubbard, R. & Waller, D.) 59–66 (SERC Daresbury Laboratory, UK, 1994).
- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *X-PLOR version 3.1—A system for X-ray crystallography and NMR* (Yale University Press, New Haven and London, 1992).
- Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923 (1976).
- Kraulis, P. J. MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Bacon, D. J. & Anderson, W. F. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220 (1988).
- Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Acknowledgements. Supported by the NIH (D.C.R. and J.B.H.), the NSF (J.B.H.) and by Deutsche Forschungsgemeinschaft postdoctoral fellowships (to C.K. and H.S.). X-PLOR calculations were performed on the CRAY C90 at the San Diego Supercomputer Center, supported by the NSF.

Correspondence and request for materials should be addressed to J.B.H. (e-mail: howar001@maroon.tc.umn.edu) or D.C.R. (e-mail: rees@citray.caltech.edu). Coordinates will be deposited in the Brookhaven Protein Data Base (ID 1N2C).

Petrological evidence for shock melting of carbonates in the martian meteorite ALH84001

Edward R. D. Scott, Akira Yamaguchi & Alexander N. Krot

Hawaii Institute of Geophysics and Planetology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

The meteorite ALH84001—a shocked igneous rock of probable martian origin—contains chemically and isotopically heterogeneous carbonate globules^{1–8}, associated with which are organic and inorganic structures that have been interpreted⁷ as possible fossil remains of ancient martian biota. A critical assumption underlying this suggestion is that the carbonates formed from low-temperature fluids penetrating the cracks and voids of the host rock³. Here we report petrological studies of ALH84001 which investigate the effects of shock on the various mineralogical components of the rock. We find that carbonate, plagioclase and silica were melted and partly redistributed by the same shock event responsible for the intense local crushing^{1,2} of pyroxene in the meteorite. Texture and compositional data show that, during the period of shock decompression, monomineralic melts were injected into pyroxene fractures that were subsequently cooled and resealed within seconds. Our results therefore suggest that the carbonates in ALH84001 could not have formed at low temperatures, but instead crystallized from shock-melted material; this conclusion weakens significantly the arguments that these carbonates could host the fossilized remnants of biogenic activity.

The martian meteorite ALH84001 is unique in containing 1 vol.% Mg-Fe-Ca-Mn carbonates, which are dispersed in an orthopyroxenite rock with a few vol.% of plagioclase glass, chromite and silica^{1,2}. Intensive scrutiny of the carbonates has produced many models for their origin but none satisfies all constraints. Some models invoke reaction of fluids with plagioclase, olivine or pyroxene at high^{1,6} or low² temperatures. Others envisage low-temperature deposition in cracks and pores^{3,7}. Given certain assumptions^{4,5}, carbon and oxygen isotopic data appear consistent with deposition at low temperatures (0–80 °C)^{3,7}. Carbonate formation by reaction of silicates with an impact-produced pulse of CO₂-rich fluid above 650 °C (ref. 6) satisfies some but not all constraints^{8,9}. Low-temperature, disequilibrium precipitation from fluids is currently the favoured mechanism.

Grains of plagioclase-rich glass 2–50 µm in size are widely distributed in the crushed zones (or granular bands²), on boundaries between large pyroxenes, and in fractures inside pyroxene crystals (Fig. 1a–c). This glass formed by shock and was previously identified as maskelynite^{1–3,6}, which forms from plagioclase without any change in the shape and composition, probably by melting¹⁰. Because the shapes of all plagioclase glasses and the compositions of many appear to have been modified by shock, use of the term ‘maskelynite’ is inappropriate and misleading.

Narrow veins of plagioclase glass in pyroxene were formed by injection of impact melt into fractures (Fig. 1a). Elsewhere, rounded elongated blebs are aligned on what appear to be healed fractures in pyroxene crystals (Fig. 1b). In crushed zones, relatively unfractured plagioclase glass locally encloses pyroxene grains and fills fractures indicating that the glass formed after crushing (Fig. 1c). Large glass grains elsewhere show textural evidence of melting and flow at their edges. We found one vesiculated grain.

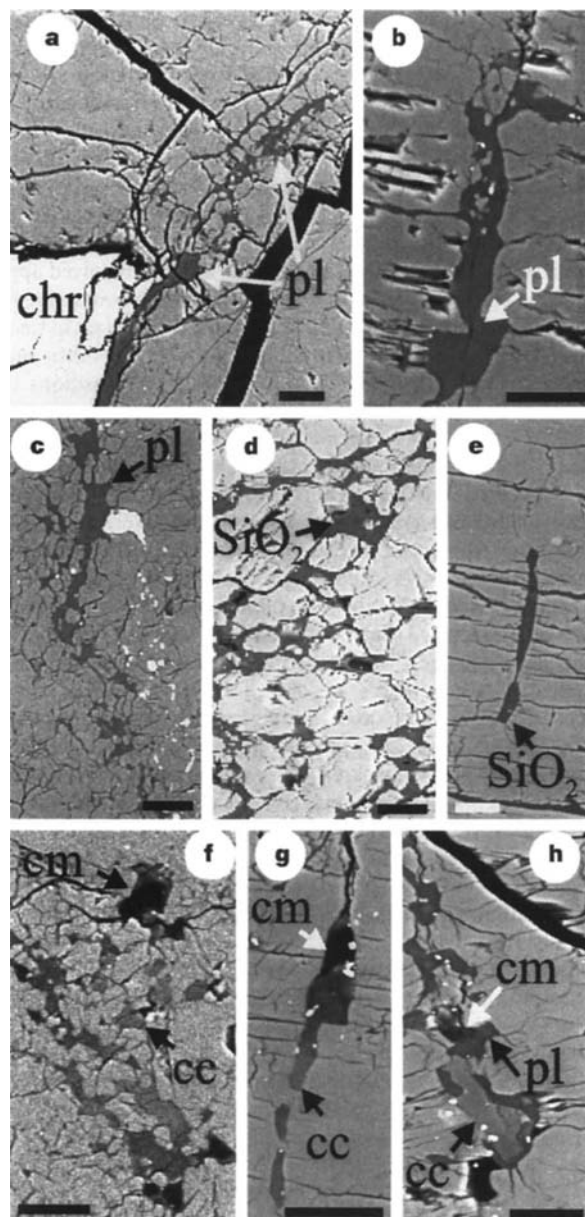


Figure 1 Back-scattered electron images of ALH84001 (section number 53) showing dark grains of plagioclase glass (pl), silica glass (SiO₂), and carbonate (cc, cm) in fractured and crushed orthopyroxene (light colour). **a**, Irregular fractures in a pyroxene crystal filled with plagioclase glass; fractures are concentrated at the tip of an enclosed chromite (chr). **b**, Elongated grain of plagioclase glass that apparently formed when shock melt was injected up an irregular fracture in a pyroxene crystal; some melt also penetrated sideways into adjacent parallel fractures. **c**, Region of crushed zone containing plagioclase glass between crushed pyroxenes. **d**, Part of a 400 × 600 µm area of crushed and fractured pyroxene containing numerous 1–10 µm silica grains enclosing small rounded pyroxenes and filling pyroxene fractures. **e**, An elongated, pinched grain of silica in pyroxene; it and nearby inclusions are aligned on parallel, healed fractures. **f**, Triangular-shaped carbonate grain in a crushed zone enclosing and surrounded by crushed pyroxene. **g**, Elongated, pinched carbonate grain inside a pyroxene crystal located on a healed crack that also contains an elongated inclusion of plagioclase glass. In both **f** and **g** the carbonate grains are zoned with a lighter, relatively Ca-rich region near the centre (cc) and darker Mg-rich regions (cm) at the top and bottom. **h**, Elongated mixed grain of plagioclase glass and carbonate in a healed fracture in pyroxene. The morphology of the carbonate, plagioclase and silica inclusions varies according to their local abundance. Narrow veins and small isolated inclusions are irregular (**a**, **c**) whereas larger inclusions (**b**, **d**, **f**) tend to have subhedral and rounded boundaries. Broad black irregular lines visible in **a** and **h** are more recent, open fractures. Scale bars are all 20 µm.

Our electron-probe analyses of plagioclase glass with a rastered beam confirm that some grains show diverse compositions and deviate grossly from plagioclase stoichiometry¹. We found some K-rich regions ($K/(Na + K + Ca) = 15\text{--}35\%$) with deviations of 3–8% in total $Ca + Na + K$ from plagioclase stoichiometry (Table 1, column 2). X-ray scanning maps of these regions reveal heterogeneous K, Na and Ca distributions and rare, micrometre-sized Si-rich grains, presumably SiO_2 -rich glasses¹¹. These effects are clearly not due to Na loss during analysis¹ and probably reflect minor shock volatilization. Pyroxene does not appear to have dissolved appreciably in the plagioclase melt, but chromium enrichments were found in one region near chromite (Table 1, column 3). Glasses in fractures are too small for rastered-beam analysis, but focused-beam analyses (after correction for Na losses) show similar compositions to the large grains (Table 1, columns 4, 5). For comparison, we also analysed maskelynite in the martian meteorite Shergotty¹² and in the Taiban chondrite¹³ using a rastered beam (Table 1, column 6), and confirmed that many ALH84001 glasses are much more heterogeneous and less stoichiometric than maskelynite.

Mobilized plagioclase melts have been reported in lunar samples and experimentally shocked rocks¹⁴ but not in strongly shocked chondrites¹³, which contain mixed-mineral and whole-rock melts. However, in fractured silicates in the Rose City¹³ meteorite we found plagioclase-rich melt veins that contain normative plagioclase contents of >90% (Table 1, column 7). We infer that they and other heterogeneous¹⁵ and K-rich¹⁶ plagioclase-like grains in shocked meteorites are impact-melt glasses like those in ALH84001.

Partly because plagioclase glass was identified as maskelynite, Treiman² suggested that the glass and crushed zones formed in separate shocks. He inferred that the zones recrystallized during metamorphism and that plagioclase glasses formed subsequently. However, micrometre-wide, plagioclase-glass films in fractures (Fig. 1a, c) would not survive such metamorphism. Crushed zones were probably heated and welded by friction during crushing, as in pseudotachylites. Because shock veins can crystallize high-pressure minerals like majorite-pyroxene garnet from melt¹⁷, we infer that one impact could have formed the crushed zones and plagioclase glass during shock decompression.

Silica grains, like plagioclase glasses, occur in healed pyroxene

fractures and in crushed zones (Fig. 1d, e) but are much less common. The absence in Raman spectra of features from crystalline silica suggests that these grains are glasses. Silica grains are rather pure, like plagioclase glasses (Table 1) and probably formed by shock-melting of interstitial grains and dispersal during crushing. Although several shocks could be responsible, one shock is plausible. L chondrites suffered one impact which fractured silicates, injected molten metallic Fe, Ni and troilite into fractures and then resealed the fractures during shock decompression^{13,15,18}.

Shock pressures of >45–50 GPa are needed, at least locally, to form and mobilize silica and plagioclase melts¹⁹ assuming that the rock was initially non-porous and cool (<100 °C). However, in our section pyroxenes outside crushed zones show little mosaicism suggesting that equilibrium shock pressures were lower^{19,20} and that the rock may have been warmer initially²⁰. Shocks of 25–45 GPa would have produced post-shock temperature increases in pyroxene of ~200–500 °C as pyroxene is less compressible than plagioclase and silica¹³.

Carbonate forms globules in plagioclase glass in uncrushed regions, grains that poikilitically enclose pyroxene in crushed zones, and flattened globules and smaller grains in pyroxene fractures^{1,2,7}. Globules are absent in our section but we studied many carbonate grains 2–200 µm in size in fractures and crushed zones (Fig. 1f, g). Small carbonate inclusions are numerous in pyroxene fractures¹, and are aligned and distributed along irregular and planar cracks that appear to have been resealed at high temperature after carbonate was introduced. The carbonate inclusions are not typically parallel-sided like fractures that form at low temperatures. Instead inclusions, especially larger ones, tend have rounded or slightly curved and cusped edges (Fig. 1g) indicating high temperatures during or after emplacement.

Carbonates in crushed zones have some textural characteristics of plagioclase and silica glasses (Fig. 1c, d, f). Carbonates enclose pyroxenes and fill fractures between broken pyroxene crystals. Large grains in crushed zones and fractures tend to have rounded and subhedral shapes whereas isolated, narrow inclusions tend to be angular (Fig. 1). This indicates hot emplacement of carbonate, not later metamorphism. In slower-cooled regions, the surface energy of grains was reduced by minor movement of boundaries.

The morphological similarities between carbonate, plagioclase and silica inclusions in healed fractures (Fig. 1b, e, g) and in crushed regions (Fig. 1c, d, f), and the presence of cracks containing both mixed (Fig. 1h) and separate inclusions of carbonate and glass, suggest that carbonate, plagioclase and silica were shock-melted and mobilized together. Rapid cooling of tiny melt volumes prevented mixing. We failed to observe mineralogical and textural evidence for formation of carbonate and silica by reaction of olivine, pyroxene or plagioclase with CO_2 -rich fluids^{1,2,6}. Olivine and silica are not generally associated with carbonate, and plagioclase is not rimmed by carbonate in mixed grains (Fig. 1h).

On a $MgCO_3$ – $CaCO_3$ – $FeCO_3$ plot, all carbonate occurrences show the same trend (Fig. 2) indicating a common origin. We infer that all carbonates crystallized from impact melts (or fluids). Cogent arguments⁸ against the impact model of Harvey and McSweeney⁶ do not apply to our model. In the former, CO_2 -rich fluids infiltrate rocks at depth (>50 km, ref. 8), whereas in the latter pre-existing carbonates melt under shock pressure in seconds. Carbonates in cracks cannot be terrestrial contaminants, as Jull *et al.*⁴ surmised. Discrepant ages for carbonate and plagioclase and heterogeneous carbon isotopic compositions require further study.

The extremely low viscosity of carbonate melts even under pressure²¹ accounts for both the high abundance of carbonate in large fractured pyroxenes relative to plagioclase and silica, and the difficulty in quenching carbonate melts to glass. The globular shape of carbonates in plagioclase glass^{1,2} is typical of carbonates crystallizing from synthetic carbonate–silicate melts²². Experiments sug-

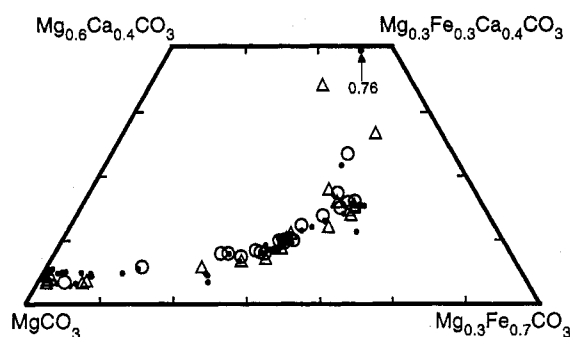


Figure 2 Molar compositions of carbonates in ALH84001 projected onto a $MgCO_3$ – $FeCO_3$ – $CaCO_3$ plot. Our electron-probe data for grains in fractures in pyroxene (triangles) and grains that poikilitically enclose pyroxene in the crushed zones (dots) are not significantly different from each other or from the data of Mittlefehdt¹ for the large globules in plagioclase glass (circles) and Valley *et al.*⁸ for flattened globules. Each grain is zoned from $Mg_{0.93}Fe_{0.03}Ca_{0.04}CO_3$ to around $Mg_{0.5}Fe_{0.35}Ca_{0.15}CO_3$ and locally to more calcic compositions. Grains with $Ca_{>0.3}$ are rare in all locations (see also ref. 6). The most calcic region that we analysed has a composition $Mg_{0.15}Fe_{0.09}Ca_{0.76}CO_3$ and is marked with an arrow. $MnCO_3$ concentrations are typically 0–4 mol% and are positively correlated with $CaCO_3$. Like Mittlefehdt¹ we found 0.05–0.2 wt% Na_2O and SiO_2 . Treiman² showed that poikilitic and globular carbonates have similar compositions and origins; we infer that all types of carbonates formed in the same way.

Table 1 Compositions of plagioclase-rich glass, maskelynite and silica

	1	2	3	4*	5*	6	7	8*
Analyses	8	1	5	6	6	5	8	6
SiO ₂	59.9	63.9	61.2	61.3	59.6	56.9	64.1	99.2
TiO ₂	0.03	0.04	0.09	0.05	0.05	0.03	0.15	0.02
Al ₂ O ₃	25.3	21.1	23.4	25.8	25.0	26.3	20.1	0.03
Cr ₂ O ₃	0.12	0.07	1.0	0.07	0.07	<0.02	0.20	<0.02
FeO	0.45	0.41	0.92	0.65	0.63	0.69	1.8	0.64
MnO	<0.02	<0.02	0.04	0.04	0.04	<0.02	<0.02	0.02
MgO	<0.02	0.04	<0.02	0.02	0.02	<0.02	2.0	0.02
CaO	6.8	3.0	5.2	6.6	6.4	9.3	3.0	0.02
Na ₂ O	7.3	5.6	7.6	5.7	7.4	5.8	9.2	<0.03
K ₂ O	0.62	5.0	1.1	0.78	0.75	0.41	0.37	n.a.
Total	100.6	99.1	100.7	101.1	100.0	99.6	101.0	99.9
Structural formulae based on 8 O								
Si	2.66	2.88	2.72	2.69	2.67	2.57	2.83	
Ti	0.001	0.001	0.003	0.002	0.002	0.001	0.005	
Al	1.32	1.12	1.23	1.34	1.32	1.40	1.05	
Cr	0.004	0.002	0.037	0.002	0.002	0.000	0.007	
Fe	0.017	0.015	0.034	0.024	0.023	0.026	0.068	
Mn	0.001	0.000	0.002	0.002	0.001	0.001	0.001	
Mg	0.000	0.003	0.001	0.001	0.001	0.001	0.133	
Ca	0.325	0.145	0.250	0.312	0.309	0.449	0.144	
Na	0.632	0.487	0.656	0.487	0.641	0.510	0.787	
K	0.035	0.287	0.062	0.043	0.043	0.024	0.021	
Total	5.004	4.942	4.998	4.901	5.012	4.984	5.042	
Na + K + Ca	0.992	0.919	0.968	0.842	0.993	0.983	0.951	
Ab	63.7	53.0	67.7	57.9	64.6	51.9	82.7	
Or	3.5	31.2	6.4	5.4	4.3	2.4	2.2	

Compositions of plagioclase-rich glass from ALH84001 are given in columns 1–5: 1, glasses with plagioclase stoichiometry; 2, K-rich region; 3, Cr-rich region; 4, small plagioclase grains in Fig. 1a–c, uncorrected, and 5, corrected, for Na loss. Column 6, maskelynite from Shergotty. Column 7, plagioclase-rich glass from Rose City. Column 8, silica from ALH84001. Ab (per cent albite) = $\text{Na}/(\text{Na} + \text{K} + \text{Ca}) \times 100$; Or (per cent orthoclase) = $\text{K}/(\text{Na} + \text{K} + \text{Ca}) \times 100$.

* Focused beam analyses; other analyses obtained by rastering over an area of $>30 \mu\text{m}^2$. All analyses obtained with a beam current of 5 nA.

gest that carbonates are readily heated and mobilized by shock pressures of 10–60 GPa, but specific pressures and post-shock conditions needed to cause vaporization, melting, devolatilization and recrystallization are poorly constrained (see for example, refs 23, 24). Small carbonate globules have been observed in some terrestrial impact glasses²⁵.

The CaCO_3 – MgCO_3 phase diagram suggests that crystallization of magnesite from a melt near equilibrium would require pressures of at least 3 GPa; calcite requires²⁶ 0.3 GPa. Devolatilization of carbonate may have been prevented by crystallization during shock decompression. Submicrometre pores in magnetite-rich carbonate regions⁷ may result from minor devolatilization during crystallization rather than subsequent dissolution. A melt with the mean Ca/Mg ratio of the carbonates in Fig. 2 should crystallize grains with Mg-rich cores and Ca-rich rims, contrary to what is observed^{1–3,6,7}. However, carbonate melts readily undercool and form disequilibrium phases²⁷. We speculate that the martian carbonates may have crystallized initially at 0.3–3 GPa. Carbonate spherules that crystallized rapidly in peridotite xenoliths from Spitzbergen²⁸ show zoning patterns that qualitatively resemble those in ALH84001.

We do not know the origin of the carbonates that existed in ALH84001 before shock melting and hope to address this issue when sample allocations are resumed. Nevertheless, our evidence for shock mobilization of carbonates argues strongly against the key assumption of McKay *et al.*⁷ that the existing carbonates formed at low temperatures, and their conclusion that the carbonates formed biogenically. Submicrometre magnetites and sulphides in carbonates⁷ (also proposed as possible evidence for biogenic activity) may have formed during crystallization of carbonate melts. Organic compounds of possible biogenic origin that are associated with martian carbonate^{7,29} could conceivably have survived shocks. Alternatively, they may have formed during post-shock cooling³⁰ or been deposited subsequently. □

Received 18 November 1996; accepted 8 April 1997.

- Mittlefehldt, D. W. ALH84001, a cumulate orthopyroxenite member of the martian meteorite clan. *Meteoritics* **29**, 214–221 (1994).
- Treiman, A. H. A petrographic history of martian meteorite ALH84001: two shocks and an ancient age. *Meteoritics* **30**, 294–302 (1995).
- Romanek, C. S. *et al.* Record of fluid-rock interactions on Mars from the meteorite ALH84001. *Nature* **372**, 655–657 (1994).
- Jull, A. J. T., Eastoe, C. J., Xue, S. & Herzog, G. F. Isotopic composition of carbonates in the SNC meteorites Allan Hills 84001 and Nakhla. *Meteoritics* **30**, 311–318 (1995).
- Jull, A. J. T., Eastoe, C. J. & Clout, S. Isotopic composition of carbonates in the SNC meteorites, Allan Hills 84001 and Zagami. *J. Geophys. Res.* **102**, 1663–1669 (1997).
- Harvey, R. P. & McSweeney, H. Y. Jr A possible high-temperature origin for the carbonates in the martian meteorite ALH84001. *Nature* **382**, 49–51 (1996).
- McKay, D. S. *et al.* Search for past life on Mars: possible relic biogenic activity in martian meteorite ALH84001. *Science* **273**, 924–930 (1996).
- Valley, J. W. *et al.* Low-temperature carbonate concretions in the martian meteorite ALH84001: evidence from stable isotopes and mineralogy. *Science* **275**, 1633–1638 (1997).
- Kirschvink, J. L., Maine, A. T. & Vali, H. Paleomagnetic evidence of a low-temperature origin of carbonate in the martian meteorite ALH84001. *Science* **275**, 1629–1633 (1997).
- Stöfler, D. Glasses formed by hypervelocity impact. *J. Non-cryst. Solids* **67**, 465–502 (1984).
- Thomas, K. L. *et al.* Microanalysis of unique fine-grained minerals in the martian meteorite ALH84001. *Lunar Planet. Sci.* **27**, 1327–1328 (1996).
- Stöfler, D. *et al.* Shock metamorphism and petrography of the Shergotty achondrite. *Geochim. Cosmochim. Acta* **50**, 889–903 (1986).
- Stöfler, D., Keil, K. & Scott, E. R. D. Shock metamorphism of ordinary chondrites. *Geochim. Cosmochim. Acta* **55**, 3845–3867 (1991).
- Schaal, R. B. & Hörz, F. Shock metamorphism of lunar and terrestrial basalts. *Proc. Lunar Sci. Conf.* **8**, 1697–1729 (1977).
- Rubin, A. E. A shock-metamorphic model for silicate darkening and compositionally variable plagioclase in CK and ordinary chondrites. *Geochim. Cosmochim. Acta* **56**, 1705–1714 (1992).
- Bischoff, A., Rubin, A. E., Keil, K. & Stöfler, D. Lithification of gas-rich chondrite regolith breccias by grain boundary and localized shock melting. *Earth Planet. Sci. Lett.* **66**, 1–10 (1983).
- Chen, M., Sharp, T. G., El Goresy, A., Wopenka, B. & Xie, X. The majorite-pyroxene + magnesio-wüstite assemblage: constraints on the history of shock veins in chondrites. *Science* **271**, 1570–1573 (1996).
- Joreau, P., Leroux, H. & Doukhan, J.-C. A transmission electron microscope investigation of shock metamorphism in olivine of the Ilafegh 013 chondrite. *Meteorit. Planet. Sci.* **32**, 309–316 (1997).
- Stöfler, D., Bischoff, A., Buchwald, V. F. & Rubin, A. E. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 165–202 (Univ. Arizona Press, Tucson, 1988).
- Schmitt, R. T., Deutsch, A. & Stöfler, D. Shock recovery experiments with the H6 chondrite Kernouvé at preshock temperatures of 293 and 920 K. *Meteoritics* **29**, 529–530 (1994).
- Dobson, D. P. *et al.* In-situ measurement of viscosity and density of carbonate melts at high pressure. *Earth Planet. Sci. Lett.* **143**, 207–215 (1996).
- Lee, W.-J. & Wyllie, P. J. Liquid immiscibility in the join $\text{NaAlSi}_3\text{O}_8$ – CaCO_3 to 2.5 GPa and the origin of calcic-carbonate magmas. *J. Petrol.* **37**, 1125–1152 (1996).
- Martinez, I. *et al.* Shock recovery experiments on dolomite and thermodynamical calculations of impact induced decarbonation. *J. Geophys. Res.* **100**, 15465–15476 (1995).
- Lange, M. A. & Ahrens, T. J. Shock-induced CO_2 loss from CaCO_3 ; implications for early planetary atmospheres. *Earth Planet. Sci. Lett.* **77**, 409–418 (1986).
- Dence, M. R., von Engelhardt, W., Plant, A. G. & Walter, L. S. Implications of fluid immiscibility in glass from West Clearwater Lake Impact Crater, Quebec, Canada. *Contrib. Mineral. Petrol.* **46**, 81–97 (1974).
- Wyllie, P. J. in *Carbonatites: Genesis and Evolution* (ed. Bell, K.) 500–545 (Unwin Hyman, London, 1989).
- Irving, A. J. & Wyllie, P. J. Subsolidus and melting relationships for calcite, magnesite and the join CaCO_3 – MgCO_3 to 36 kb. *Geochim. Cosmochim. Acta* **39**, 35–53 (1975).
- Ionov, D. A. *et al.* Carbonated peridotite xenoliths from Spitzbergen: implications for trace element signature of mantle carbonate metasomatism. *Earth Planet. Sci. Lett.* **119**, 283–297 (1993).
- Wright, I. P., Grady, M. M. & Pillinger, C. T. Organic materials in a martian meteorite. *Nature* **340**, 220–222 (1989).
- Semenenko, V. P. & Golovka, N. V. Shock-induced black veins and organic compounds in ordinary chondrites. *Geochim. Cosmochim. Acta* **58**, 1525–1535 (1994).

Acknowledgements. We thank A. H. Treiman, G. J. Taylor, L. B. Browning, A. J. T. Jull, P. H. Warren and many other colleagues for helpful discussions, T. F. Cooney and S. K. Sharma for assistance with Raman spectroscopy, A. Rubin for helpful comments, and R. Score and colleagues for field work in Antarctica. This work was partly supported by NASA.

Correspondence should be addressed to E.R.D.S. (e-mail: escott@pgd.hawaii.edu).

Synthesis and structural characterization of an Al₇₇ cluster

A. Ecker*, E. Weckert† & H. Schnöckel*

* Institut für Anorganische Chemie, † Institut für Kristallographie, Universität Karlsruhe, Engesserstrasse Geb. 30.45, D-76128 Karlsruhe, Germany

Molecular species containing metal–metal bonds have been known for a long time¹. But these have generally involved transition metals; metal–metal bonds between main-group elements were discovered only more recently². Within the past decade, polyhedral Al₄, Al₁₂ and Ga₄ species have all been characterized by

X-ray diffraction^{3,4}, and the chemistry of low-valent polynuclear Al species in particular has flourished owing to the development of condensation techniques for preparing them by disproportionation of aluminium monohalides⁴. By using this approach, we now report the identification and characterization of an Al₇₇ cluster, which is dramatically larger than any seen hitherto and can be considered an intermediate species to the formation of the bulk metal. X-ray diffraction shows that the cluster has a central Al atom surrounded by three concentric polyhedral shells containing 12, 44 and 20 Al atoms. This Al₇₇ core is stabilized by 20 organic ligands, preventing the formation of the bulk metal. Studies of such clusters should provide insights into the crossover between molecular species and the bulk metal for main-group elements.

In the experiments described here our aim was to prepare polyhedral Al(I) compounds containing more than four Al atoms using N(SiMe₃)₂ ligands for protection. A solution of aluminium(I) iodide⁵ in a solvent mixture of toluene and diethyl ether was added to a solution of base-free LiN(SiMe₃)₂ at -78 °C. The temperature was allowed to rise slowly to 5 °C and finally, for almost half an hour,

to 60 °C. The solution, originally light in colour, darkened markedly and subsequently yielded black crystals which are stable at room temperature although highly sensitive to attack by air and moisture. These formed on the walls of the glass reaction vessel at ~20 °C. No solid LiI or metallic Al was observed to separate.

With the help of X-ray techniques, single crystals of the black product were found to crystallize in the triclinic space group $P\bar{1}$ (ref. 6), with a unit cell volume of 8,807 Å³. The most remarkable result of the crystal structure analysis (see Methods) was the identification of a very large metal cluster within a compound conforming to the overall formula:



where ($\mu\text{-I}$) indicates a bridging iodine atom. The Al atoms in the cluster anion $[\text{Al}_{77}\text{R}_{20}]^{2-}$, where R is N(SiMe₃)₂ (Fig. 1), correspond formally to an average oxidation number of +0.23. Starting with the transient species Al(I)R, where R is N(SiMe₃)₂, the Al₇₇ cluster is therefore revealed an intermediate species trapped during the various disproportionation reaction which precede the formation of the pure metallic state, oxidation number zero.

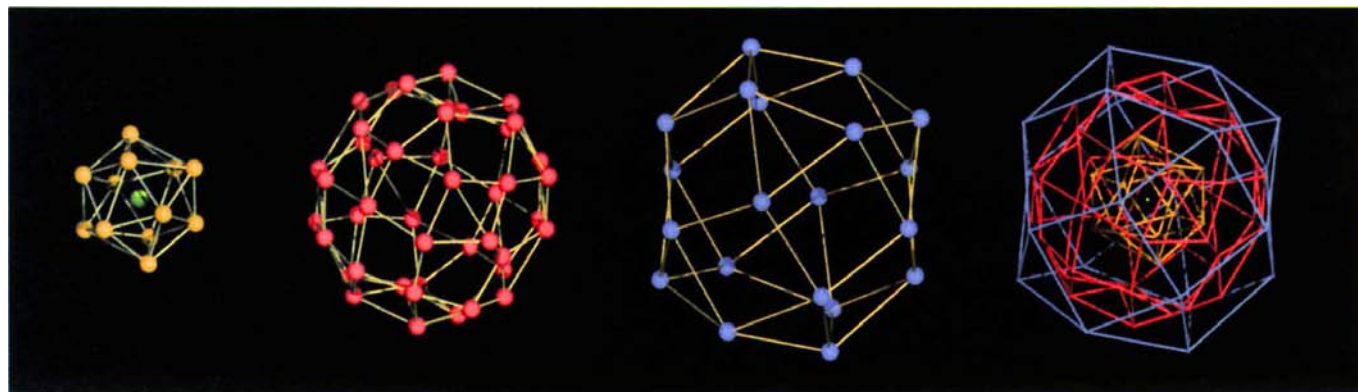


Figure 1 Separate (ball and stick models) and combined (stick model) representations of the different cluster shells of Al₇₇: 1 (central atom) + 12 (orange, distorted icosahedron), 44 (red), 20 (blue). The lines have only topologi-

cal significance; in the combined representation (top) yellow, orange and red lines are <350 pm; blue lines are <520 pm). The amide ligands are bound to the vertices of the blue polyhedron.

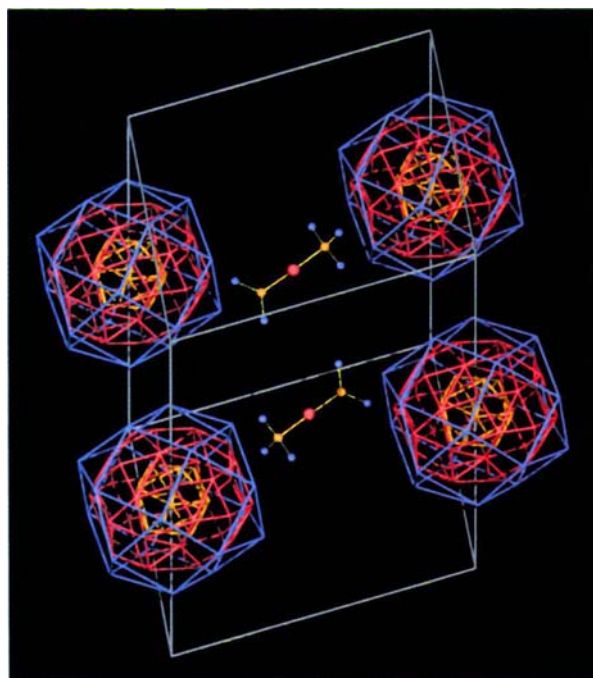


Figure 2 Triclinic unit cell with the cluster anions (stick model of the aluminium core) on four edges. The double charge of the cluster anion is deduced from the two singly charged cations (ball and stick model: red, iodine; orange, lithium; blue, oxygen of the coordinate diethyl ethers) in the centre of the unit cell. The different coordination numbers of the two Li atoms may give rise to the different Li-I distances of 263.7 and 288.6 pm. Both angles at the different Li atoms confirm this interpretation: the two I-Li-O and O-Li-O angles for the trigonal planar coordinate Li atom sum to ~360°, whereas the O-Li-O and I-Li-O angles for the other Li atom are close to the tetrahedral value.

Table 1 Coordination numbers and distances of the Al₇₇ cluster core

Coordination number*	12	10	7	4
Number of atoms	1	12	44	20
Minimum $d(\text{Al-Al})$ (pm)	267.5	264.0	256.4	256.4
Maximum $d(\text{Al-Al})$ (pm)	286.8	314.0	314.0	285.0
Average $d(\text{Al-Al})$ (pm)	276.0	283.8	276.3	268.8

Aluminium-aluminium distances ($d(\text{Al-Al})$) are given in the table if they are less than $d_{\text{metal}}(\text{Al-Al}) + 10\%$.

* Exact values are 10.3, 7.3 and 4.1.

The most interesting part of the crystal structure is the Al_{77} core, which is (to our knowledge) the largest metallic cluster species—in contrast to larger salt-like clusters such as $\text{Cu}_{146}\text{Se}_{73}$ (ref. 7)—to be characterized to date by X-ray methods. With the help of this result, it is now possible to gain a detailed insight into the geometry of homonuclear metal clusters larger than, for example, $\text{Pt}_6\text{Ni}_{38}(\text{CO})_{47}^{6-}$ (ref. 8).

The geometric arrangement of the Al_{77} cluster can be described as being composed of six shells, consisting of 12, 44 and 20 aluminium atoms, 20 nitrogen atoms, 30 silicon atoms and a 'skin' of 120 methyl groups, surrounding a central Al atom. Only this central atom is coordinated by 12 Al neighbours as in the metal itself. However, the coordination is not cuboctahedral, as in face-centred cubic aluminium metal, but rather a distorted isosahedron. Only the centre of the cluster may be considered as 'bulk-like', with the remaining 76 Al atoms having lower coordination numbers. The different Al–Al distances, as they appear in moving from the inside to the outside of the cluster, are summarized in Table 1, where we consider only Al atoms. The decreasing coordination number corresponds to the transition from the solid metal to molecular species.

The Al_{77} clusters within the triclinic crystal cell adopt the unusual simple hexagonal form of packing. Each cluster is surrounded by 8+12 neighbouring clusters with distances between the centres of 2,200 and 3,060 pm. The holes between clusters are filled with two toluene molecules of crystallization and two cations $[\text{Li}(\text{OEt})_2]_3(\mu\text{-I})\text{Li}(\text{OEt})_2]^+$ (Fig. 2). In the ESR spectrum a broad signal (even at 77 K, ΔH peak-to-peak = ~ 100 G) undoubtedly points to the presence of paramagnetic species in solution with $S \geq 1/2$. The high spin state, together with the coupling of many Al atoms, makes it difficult not only to detect but also to assign the ESR signals. Further experiments with solid samples are in progress.

The results presented here may complement the many investigations of Au_{55} clusters⁹ and related systems, which have been carried out by mass spectroscopy, electron microscopy and other methods, but not, as yet, by X-ray techniques. The unexpected formation of an Al_{77} cluster by means of new techniques, followed by an X-ray structural investigation, demonstrates that (from a geometrical point of view) there are metallic conditions with coordination number of 12—as in cubic face centred aluminium metal ($d(\text{Al}-\text{Al}) = 286$ pm)—at the centre of the cluster and molecular type bonding on the surface of the metal core.

Thus the Al_{77} cluster only represents a species at the very beginning of the transition from the molecular to bulk regime. Only in much larger clusters does there seem to be any likelihood of 'metallic' coordination spheres for atoms other than the central one. □

Methods

X-ray structure analysis. A single crystal ($0.2 \times 0.2 \times 0.1$ mm) was studied. Unit-cell dimensions: $a = 21.03(6)$ Å, $b = 21.71(6)$ Å, $c = 22.15(6)$ Å, $\alpha = 63.63(6)^\circ$, $\beta = 77.23(6)^\circ$, $\gamma = 80.17(7)^\circ$; $V = 8807(40)$ Å³; $Z = 1$; $M = 6,492.48$ g mol⁻¹; $\rho_{\text{calc}} = 1.224$ g cm⁻³; $F(000) = 3,297$. Equipment: Siemens rotating anode (0.3×3 mm, 5.0 kW) $\lambda(\text{MoK}\alpha) = 0.71073$ Å, $\mu(\text{MoK}\alpha) = 0.542$ mm⁻¹; MarResearch IPDS diffractometer (180 mm). Data processing software: XDS (ref. 10), 63,649 reflections measured, 30,092 independent ($R_{\text{int}} = 0.0665$), 18,810 observed ($F > 4\sigma(F)$). Structure determination/refinement software: SHELXS-86 (ref. 11), SHELXL-93 (G. M. Sheldrick, manuscript in preparation), $R1/wR2(F > 4\sigma(F)) = 0.1262/0.3348$, $F1/wR2(\text{all data}) = 0.1846/0.3838$, $\text{GOF}(F^2) = 1.026$. The cluster anion was refined anisotropically, whereas the cation was refined isotropically. Difficulties arose in describing the solvent molecules correctly (toluene). Further details of the crystal structure investigation may be obtained from the Fachinformationszentrum Karlsruhe, D-76344 Eggenstein-Leopoldshafen (Germany), on quoting the deposition number CSD 406760.

Received 22 January; accepted 8 April 1997.

1. Wilson, F. C. & Shoemaker, D. P. X-ray crystallographic data for metal cyclopentadienyl carbonyl compounds. *Naturwissenschaften* **43**, 57–58 (1956).

- Uhl, W. Tetrakis[*bis*(trimethylsilyl)methyl]dialan(4), eine Verbindung mit Aluminium–Aluminium-Bindung. *Z. Naturforsch. B* **43**, 1113–1118 (1988).
- Uhl, W. Elementorganische Verbindungen mit Al–Al-, Ga–Ga- und In–In-Bindung. *Angew. Chem.* **105**, 1449–1461 (1993).
- Dohmeier, C., Loos, D. & Schnöckel, H. Aluminium(1) and gallium(1) compounds: syntheses, structures and reactions. *Angew. Chem. Int. Ed Engl.* **35**, 129–149 (1996).
- Ecker, A. & Schnöckel, H. Donorstabilisiertes Aluminium(I)-iodid. *Z. Anorg. Allg. Chem.* **622**, 149–152 (1996).
- Hahn, T. (ed.) *International Tables for Crystallography* Vol. A 4th edn. (Kluwer Academic, Dordrecht, 1992).
- Krautscheid, H., Fenske, D., Baum, G. & Semmelmann, M. Ein neuer Kupferselenidcluster mit PPh₃ als Ligand: $[\text{Cu}_{146}\text{Se}_{73}(\text{PPh}_3)_{120}]$. *Angew. Chem.* **105**, 1364–1367 (1993).
- Cerioti, A. et al. Synthese und Struktur der Ionen $[\text{Ni}_{38}\text{Pt}_6(\text{CO})_{48}\text{H}_6-n]^{n-}$ ($n = 4, 5$); Ni–Pt-Cluster als molekulare Modelle für "Cherry"-Kristallite. *Angew. Chem.* **97**, 708–710 (1985).
- Schmid, G. (ed.) *Clusters and Colloids, From Theory to Applications* (VCH, Weinheim, 1994).
- Kabsch, W. Evaluation of single-crystal-X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924 (1988).
- Sheldrick, G. M. Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. A* **46**, 467–473 (1990).

Acknowledgements. We thank P. Kasai (IBM Almaden Research Center) for discussions, and VAW aluminium AG (Bonn, Germany) for contributing to the cost of colour figures. This work was supported by the DFG.

Correspondence and requests for materials should be addressed to H.S.

Residue-based control of helix shape in β -peptide oligomers

Daniel H. Appella*, Laurie A. Christianson*, Daniel A. Klein*, Douglas R. Powell*, Xiaolin Huang†, Joseph J. Barchi Jr† & Samuel H. Gellman*

* Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA

† Laboratory of Medicinal Chemistry, Division of Basic Sciences, National Cancer Institute, Bethesda, Maryland 20892, USA

Proteins and RNA are unique among known polymers in their ability to adopt compact and well-defined folding patterns. These two biopolymers can perform complex chemical operations such as catalysis and highly selective recognition, and these functions are linked to folding in that the creation of an active site requires proper juxtaposition of reactive groups. So the development of new types of polymeric backbones with well-defined and predictable folding propensities ('foldamers') might lead to molecules with useful functions^{1,2}. The first step in foldamer development is to identify synthetic oligomers with specific secondary structural preferences^{3–13}. Whereas α -amino acids can adopt the well-known α -helical motif of proteins, it was shown recently^{11–13} that β -peptides³ constructed from carefully chosen β -amino acids can adopt a different, stable helical conformation defined by interwoven 14-membered-ring hydrogen bonds (a 14-helix; Fig. 1a). Here we report that β -amino acids can also be used to design β -peptides with a very different secondary structure, a 12-helix (Fig. 1a). This demonstrates that by altering the nature of β -peptide residues, one can exert rational control over the secondary structure.

We have found previously that β -peptides derived from a rigidified β -amino acid, *trans*-2-aminocyclohexanecarboxylic acid (*trans*-ACHC), adopt the 14-helix¹³. Here we use an alternatively rigidified β -amino acid, *trans*-2-aminocyclopentanecarboxylic acid (*trans*-ACPC). Chemical structures are shown in Fig. 2. Our examination of *trans*-ACPC oligomers was motivated by the results of molecular mechanics and dynamics simulations. As previously described, computational comparisons involving eight alternative cycloalkane-based β -amino acids and six alternative β -peptide helices suggested that the combination of *trans*-ACHC and the 14-helix conformation would generate the most stable β -peptide secondary structure¹³. The *trans*-ACPC/12-helix combination was

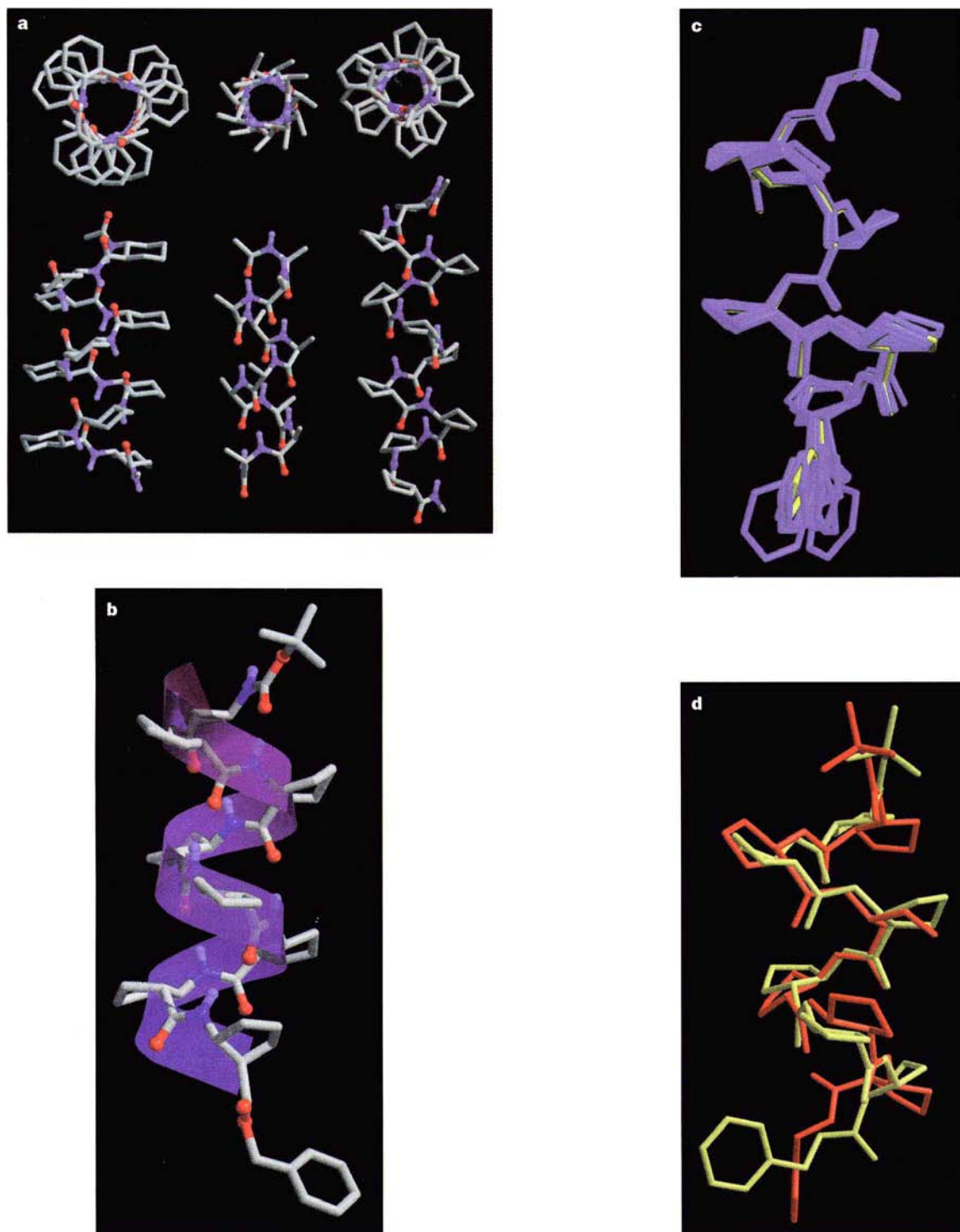


Figure 1 **a**, Comparison, left-to-right, among an idealized 14-helix ((*R,R*)-*trans*-ACHC decamer), an idealized α -helix (L-alanine decamer) and an idealized 12-helix ((*R,R*)-*trans*-ACPC decamer), in views along and perpendicular to the helix axis (N terminus is at top in latter)^{27–29}. **b**, Octamer **1** in the solid state; N terminus at top. Crystallographic data: $C_{60}H_{88}N_8O_{11} \cdot 1.85 (CH_3OH)(H_2O)$, space group $P2_12_12_1$, with $a = 20.8664(3) \text{ \AA}$, $b = 30.5453(3) \text{ \AA}$, $c = 10.1044(2) \text{ \AA}$, unit cell volume $6,440.5(2) \text{ \AA}^3$, $R(F) = 12.60\%$ for 5,980 observed ($F > 4\sigma(F)$) data. **c**, Conformations of hexamer **2** in pyridine- d_5 (N terminus at top), based on NOE-restrained molecular dynamics simulations conducted with the program Quanta using the CHARMM 23 force field³⁰. A set of 50 initial structures was generated by unrestrained dynamics at 1,000 K. Each structure was then subjected to dynamic annealing in which 50 NOEs unambiguously assigned in pyridine- d_5 served as distance restraints. The 20 resulting structures of lowest energy and fewest

distance violations were subjected to a final dynamic annealing with additional restraints, including the four 12-membered hydrogen bonds and six dihedral restraints, from HN-C β H coupling constants. The 10 resulting conformations are superimposed, with the structure closest to the average highlighted in yellow. **d**, Superposition of the solid-state conformation of hexamer **2** (red) with the NMR-derived structure (yellow) closest to the average of the 10 conformations shown in **c** (N terminus at top). Crystallographic data for **2**: $C_{48}H_{70}N_6O_9 \cdot (CH_3OH)$, space group $P1$ with $a = 8.6599(2) \text{ \AA}$, $b = 11.6888(2) \text{ \AA}$, $c = 12.9173(2) \text{ \AA}$, $\alpha = 69.207(2)^\circ$, $\beta = 81.630(2)^\circ$, $\gamma = 88.666(2)^\circ$, unit cell volume $1,208.79(4) \text{ \AA}^3$, $R(F) = 3.55\%$ for 6,469 observed ($F > 4\sigma(F)$) data. The structures of both **1** and **2** were solved by direct methods and refined by full-matrix least squares on F^2 with anisotropic displacement parameters for all non-hydrogen atoms (SHELXTL version 5 supplied by G. M. Sheldrick).

predicted to be almost as favourable, on the basis of dynamics simulations (L.A.C., unpublished results). This latter prediction was intriguing because there is no precedent for the 12-helix in the contradictory literature on polymers constructed from optically active β -amino acids. Among these polymers, poly(α -isobutyl-L-aspartate) has been particularly intensively studied, and proposed secondary structures include 14-, 16-, 18- and 20-helices, as well as the sheet form^{14–18}. As the computational predictions regarding the *trans*-ACHC/14-helix relationship had proved to be correct¹³, we undertook to test the *trans*-ACPC/12-helix prediction.

Optically active *trans*-ACPC was prepared using published protocols^{19,20}, and β -peptides were generated through standard coupling methods. Octamer **1** displays the predicted 12-helical conformation in the solid state (Fig. 1b); all six of the possible 12-membered-ring hydrogen bonds are formed. Hexamer **2** also displays the predicted 12-helical conformation in the solid state (Fig. 1d), with all four of the possible 12-membered-ring hydrogen bonds formed. In both cases, the regular helix frays at the C terminus, perhaps because the C-terminal ester group cannot serve as a hydrogen-bond donor.

Circular dichroism (CD) data for hexamer **2** in CH₃OH (Fig. 3) suggest the adoption of a distinctive secondary structure. The 14-helical conformation in β -peptides composed of acyclic residues has a far-ultraviolet CD signature comprising a maximum at about 197 nm, a zero crossing at about 207 nm and a minimum at about 215 nm (refs 11, 12). The CD signature for **2** is clearly different: there is a maximum at about 204 nm, zero crossing at about 214 nm and minimum at about 221 nm. As the CD signature of **2** in CH₃OH does not vary significantly between 2.0 mM and 0.02 mM, it seems unlikely that aggregation occurs under these conditions.

Two-dimensional ¹H NMR data obtained for hexamer **2** in pyridine-d₅ and in CD₃OH indicate that the 12-helix is highly populated in these solvents. Resonances of all NH groups, all protons at C β of each *trans*-ACPC residue and nearly all protons

at C α of each residue were resolved in pyridine-d₅, and could be assigned by means of a combination of total correlation spectroscopy (TOCSY)²¹ and rotating frame nuclear Overhauser effect spectroscopy (ROESY)²² data. The amino terminal NH (residue 1) could be identified because it was the only one of the six amidic protons that did not show a ROESY cross-peak to C α H of another residue (this NH also showed a ROESY cross-peak to the methyls of the *t*-butoxycarbonyl group). The carboxy-terminal residue could be identified because it had the only C α H that did not show a ROESY cross-peak to a NH of another residue. Assignment of these terminal resonances allowed us to 'walk through' the remaining backbone resonances by virtue of short-range nuclear Overhauser effects (NOEs) between C α H and NH_{*i*+1}.

The secondary structure of hexamer **2** in pyridine-d₅ is defined by the long-range NOEs summarized in Table 1. NOEs between C β H_{*i*} and NH_{*i*+2}, and between C β H_{*i*} and C α H_{*i*+2}, are expected for the 12-helix. All four possible NOEs between C β H_{*i*} and NH_{*i*+2} were observed for **2**, as were two of the four C β H_{*i*} → C α H_{*i*+2} NOEs; NOEs consistent with the other two C β H_{*i*} → C α H_{*i*+2} interactions were observed but could not be unambiguously assigned because of overlap of the C α H resonances of residues 3 and 4. A NOE between NH of residue 3 and the Boc methyl groups provided further evidence for 12-helical folding. In CD₃OH, NOEs consistent with all of the long-range NOEs detected in pyridine-d₅ were observed, and there did not seem to be any additional NOEs in CD₃OH. Interpretation of the CD₃OH data was less definitive, however, because of greater overlap among proton resonances (especially for C β H of the *trans*-ACPC residues). The pattern of NOEs in CD₃OH was identical for 25 mM and 2.5 mM samples of **2**; chemical shifts and linewidths were also invariant between these two concentrations, suggesting that aggregation does not occur under these conditions.

NOE-restrained molecular dynamics simulations of hexamer **2** generated 10 low-energy conformations (Fig. 1c). Superposition of these conformers shows a high degree of order (r.m.s. deviation of backbone atoms <0.2 Å), with fraying at the C terminus. The conformer from this set that is closest to the average is superimposed on the crystal structure of **2** in Fig. 1d. The backbones of the two structures agree fairly well, although there is variation at the C terminus and in individual cyclopentane ring conformations.

These results show that short β -peptides of *trans*-ACPC have a high propensity to adopt the 12-helical folding pattern. In combination with recent reports of 14-helix formation by short oligomers of β -substituted β -amino acids^{11,12} or *trans*-ACHC¹³, our results indicate that β -peptide secondary structure can be profoundly and rationally altered by manipulating torsional preferences about the C α –C β bonds of individual residues. In contrast, only modest

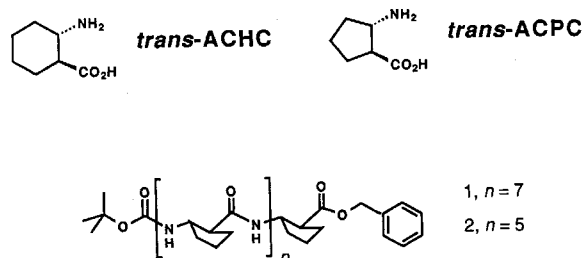


Figure 2 Chemical structures of *trans*-ACHC and *trans*-ACPC.

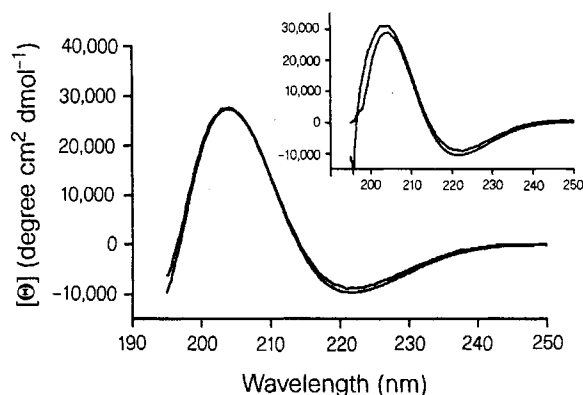


Figure 3 Circular dichroism (CD) data for hexamer **2** at 2.0 mM and 0.1 mM in CH₃OH (1 mm pathlength). Inset: CD data for **2** at 0.1 mM and 0.02 mM in CH₃OH (5 mm pathlength). Data were obtained on a Jasco J-715 instrument at 20 °C.

Table 1 Inter-residue NOEs of hexamer **2** in pyridine-d₅

Residue	H-atom	Residue	H-atom	NOE*
1	C α H	2	NH	Strong
1	C β H	2	NH	Weak
1	C β H	3	NH	Medium
1	C β H	3	C α H	Medium†
1	CH ₃ (Boc)	3	NH	Strong
1	C β H	4	NH	Weak
2	C α H	3	NH	Strong
2	C β H	4	NH	Medium
2	C β H	4	C α H	Medium†
3	C α H	4	NH	Strong
3	C β H	5	NH	Medium
3	C β H	5	C α H	Medium
4	C α H	5	NH	Strong
4	C β H	6	NH	Medium
4	C β H	6	C α H	Weak
5	C α H	6	NH	Strong
5	C β H	6	NH	Weak

* Strong, 2.3 Å; medium, 3.0 Å; weak, 4.0 Å.

† NOE present, but overlap of C α H of residues 3 and 4 prevents definite assignment.

residue-based control can be exerted over α -peptide conformation by varying substitution at C_{α} , if backbone hydrogen-bonding sites are to remain intact^{23–25}. One type of hydrogen-bonded helical secondary structure (α -helix) is predominant among peptides composed of the proteogenic α -amino acids, and the only other observed hydrogen-bonded helix (3_{10}) is relatively uncommon²⁶. The switch in helical hydrogen-bond directionality between the β -peptide 12- and 14-helices (Fig. 1a) is unprecedented among α -peptides. The predictable residue-based conformational control offered by β -peptides suggests that this class of unnatural foldamers will be well suited to molecular design efforts, such as the generation of novel tertiary structures and combinatorial searches for selective biopolymer ligands. □

Received 13 January; accepted 11 March 1997.

1. Ball, P. *Designing the Molecular World* (Princeton Univ. Press, Princeton, 1994).
2. Lehn, J.-M. *Supramolecular Chemistry: Concepts and Perspectives* (VCH, Weinheim, 1995).
3. Dado, G. P. & Gellman, S. H. Intramolecular hydrogen bonding in derivatives of β -alanine and γ -amino butyric acid: model studies for the folding of unnatural polypeptide backbones. *J. Am. Chem. Soc.* **116**, 1054–1062 (1994).
4. Hagihara, M., Anthony, N. J., Stout, T. J., Clardy, J. & Schreiber, S. L. Vinyllogous polypeptides: an alternative peptide backbone. *J. Am. Chem. Soc.* **114**, 6568–6570 (1992).
5. Eschenmoser, A. Chemistry of potentially prebiological natural products. *Origins Life* **24**, 389–423 (1994).
6. Hamuro, Y., Geib, S. J. & Hamilton, A. H. Oligoanthranilamides. Non-peptide subunits that show formation of specific secondary structure. *J. Am. Chem. Soc.* **118**, 7529–7541 (1996).
7. Gennari, C., Salom, B., Potenza, D. & Williams, A. Synthesis of sulfonamido-pseudopeptides: new chiral unnatural oligomers. *Angew. Chem. Int. Edn Engl.* **33**, 2067–2069 (1994).
8. Lokey, R. S. & Iverson, B. L. Synthetic molecules that fold into a pleated secondary structure in solution. *Nature* **375**, 303–305 (1995).
9. Smith, A. B. *et al.* De novo design, synthesis and X-ray crystal structures of pyrrolinone-based β -strand mimetics. *J. Am. Chem. Soc.* **116**, 9947–9962 (1994).
10. Nowick, J. S., Mahrus, S., Smith, E. M. & Ziller, J. W. Triurea derivatives of diethylenetriamine as potential templates for the formation of artificial β -sheets. *J. Am. Chem. Soc.* **118**, 1066–1072 (1996).
11. Seebach, D. *et al.* β -Peptides: synthesis by Arndt–Eistert homologation with concomitant peptide coupling. Structure determination by NMR and CD spectroscopy and by X-ray crystallography. Helical secondary structure of a β -hexapeptide in solution and its stability towards pepsin. *Helv. Chim. Acta* **79**, 913–941 (1996).
12. Seebach, D. *et al.* Probing the helical secondary structure of short-chain β -peptides. *Helv. Chim. Acta* **79**, 2043–2066 (1996).
13. Appella, D. H., Christianson, L. A., Karle, I. L., Powell, D. R. & Gellman, S. H. β -Peptide foldamers: robust helix formation in a new family of β -amino acid oligomers. *J. Am. Chem. Soc.* **118**, 13071–13072 (1996).
14. Yuki, H., Okamoto, Y., Taketani, Y., Tsubota, T. & Marubayashi, Y. Poly(β -amino acid)s. IV. Synthesis and conformational properties of poly(α -isobutyl-L-aspartate). *J. Polym. Sci. Polym. Chem. Edn* **16**, 2237–2251 (1978).
15. Fernández-Santín, J. M., Aymami, J., Rodríguez-Galan, A., Muñoz-Guerra, S. & Subirana, J. A. A pseudo α -helix from poly(α -isobutyl-L-aspartate), a nylon-3 derivative. *Nature* **311**, 53–54 (1984).
16. Fernández-Santín, J. M. *et al.* Helical conformations in a polyamide of the nylon-3 family. *Macromolecules* **20**, 62–68 (1987).
17. Bella, J., Alemán, C., Fernández-Santín, J. M., Alegre, C. & Subirana, J. A. Conformation of the helical polyamide poly(α -isobutyl-L-aspartate). *Macromolecules* **25**, 5225–5230 (1992).
18. López-Carrasquero, F., Alemán, C. & Muñoz-Guerra, S. Conformational analysis of helical poly(β -L-aspartate)s by IR dichroism. *Biopolymers* **36**, 263–271 (1995).
19. Herradon, B. & Seebach, D. Monoalkylation and dialkylation of derivatives of (1R, 2S)-2-hydroxycyclopentanecarboxylic acid and (1R, 2S)-2-hydroxycyclohexanecarboxylic acid via bicyclic dioxanones: selective generation of three contiguous stereogenic centers on a cyclohexane ring. *Helv. Chim. Acta* **72**, 690–714 (1989).
20. Tilley, J. W. *et al.* Analogs of Ac-CCK-7 incorporating dipeptide mimics in place of Met²⁹–Gly²⁹. *J. Med. Chem.* **35**, 3774–3783 (1992).
21. Braunschweiler, L. & Ernst, R. R. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521–528 (1983).
22. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D. & Jeanloz, R. W. Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811–813 (1984).
23. Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd Edn (Freeman, New York, 1993).
24. Xiong, H., Buckwalter, B. L., Shieh, H.-M. & Hecht, M. H. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc. Natl Acad. Sci. USA* **92**, 6349–6353 (1995).
25. Schenck, H. L., Dado, G. P. & Gellman, S. H. Redox-triggered secondary structure changes in the aggregated states of a designed methionine-rich peptide. *J. Am. Chem. Soc.* **118**, 12487–12494 (1996).
26. Barlow, D. J. & Thornton, J. M. Helix geometry in proteins. *J. Mol. Biol.* **201**, 601–619 (1988).
27. Bacon, D. J. & Anderson, W. F. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* **6**, 219–220 (1988).
28. Merrit, E. A. & Murphy, M. E. P. Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Cryst. D50*, 869–873 (1994).
29. Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950 (1991).
30. Brooks, B. R. *et al.* CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

Acknowledgements. This work was supported by the US National Science Foundation (S.H.G.) and the AIDS Targeted Antiviral Research Program for the Office of the Director of the National Institutes of Health (J.B.J. and X.H.). D.A. was supported by a Chemistry–Biology Interface Training Grant from NIGMS. L.C. was supported in part by a Graduate Fellowship from the Office of Naval Research.

Correspondence regarding the NMR-based solution structure of the hexamer should be directed to J.B.J. (e-mail: barchij@dc37a.nci.nih.gov), and correspondence regarding other issues should be addressed to S.H.G. (e-mail: gellman@chem.wisc.edu).

Modelling teleconnections between the North Atlantic and North Pacific during the Younger Dryas

Uwe Mikolajewicz*, Thomas J. Crowley†, Andreas Schiller‡ & Reinhard Voss§

* Max-Planck-Institut für Meteorologie, D-20146 Hamburg, Germany

† Department of Oceanography, Texas A&M University, College Station, Texas 77843, USA

‡ CSIRO Division of Marine Research, Hobart TAS 7001, Australia

§ Deutsches Klimarechenzentrum, D-20146 Hamburg, Germany

Evidence for a cooling event synchronous with the Younger Dryas (12,000 calendar years before present) has been found in the North Pacific Ocean north of 30°N in records of surface^{1–5} and subsurface water properties^{6,7}. These changes may be related to a temporary shut-down of North Atlantic Deep Water formation and associated surface cooling over the North Atlantic. It has remained unclear, however, whether this North Atlantic cooling was communicated to the North Pacific Ocean through the atmosphere or the ocean. Here we report results of a sensitivity experiment with a coupled ocean–atmosphere general circulation model that support a primarily atmospheric forcing of North Pacific climate variations. Changes in wind strongly affect coastal upwelling at the North American west coast, and surface cooling by the atmosphere causes better ventilation of the thermocline waters of the northeast Pacific. This effect is amplified by oceanic propagation to the Pacific of the signal arising from collapse of North Atlantic Deep Water formation. These teleconnections may also explain earlier North Pacific and western North American millennial-scale cooling events of a similar nature^{8–12}.

Previous modelling results indicate that meltwater-induced changes in subpolar North Atlantic salinity affects North Atlantic Deep Water (NADW) production (see, for example, refs 13–18). Although most of the analyses of these runs have focused on the North Atlantic and surroundings, less attention has been paid to the far-field effect of meltwater-induced NADW decreases. Generally this effect is significantly smaller than in the North Atlantic. The model¹⁸ used here shows a maximum cooling over the North Atlantic and Europe, but almost the entire Northern Hemisphere experiences a marked cooling (Fig. 1). This response is more extensive than that obtained with a different coupled model¹⁷, but the brevity of the meltwater forcing in that run (10 years) does not make the two studies strictly comparable. However, an equilibrium run¹⁹ with a rather similar model results in a response quite similar to ours. From simulations with less complete models^{15,20,21}, enhanced formation of intermediate water in the North Pacific has been reported as a consequence of the shutdown of NADW formation.

The model used here is the ECHAM3/LSG coupled ocean–atmosphere general circulation model (OAGCM) consisting of the spectral atmosphere model ECHAM3²² with a T21 resolution and 19 levels, and the LSG ocean model²³ with a horizontal resolution of 5.6° and 11 levels. The two model components are coupled by the fluxes of heat, mass and momentum, making use of the flux correction technique. Both components of the OAGCM are periodically synchronously coupled. Periods with synchronous coupling (both models are integrated quasi-simultaneously) of 15 months alternate with ocean-only periods of 48 months²⁴. This technique saves considerable amounts of computer time, while retaining the response on timescales longer than a few decades. A

more detailed description of the coupled model and its climate can be found elsewhere^{18,25}.

In a sensitivity experiment we investigated the stability of the thermohaline circulation against meltwater input into the Labrador Sea, keeping all other boundary conditions (such as topography, vegetation and ice sheets) at modern values. The prescribed triangular-shaped meltwater spike lasted for 500 years and reached a maximum of 0.625 Sv ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$) in year 250. The total length of this experiment and of an unperturbed control run is 850 model years. This experimental set-up mimics a meltwater event originating from the Laurentide ice sheet over North America. The prescribed discharge rates are roughly consistent with the estimates of the maximum rate of change of global mean sea level during the last deglaciation²⁶. But this model, like other models (for example, refs 14, 17, 20) fails to reproduce the delay of about 1,000 years between the peak of the meltwater input and the onset of the Younger Dryas.

As discussed more fully in ref. 18, the meltwater input into the Labrador Sea reduced the model's surface salinity in the northern North Atlantic, and thus formation of NADW ceased (Fig. 2a). As a result, the thermohaline circulation of the Atlantic was reversed, accompanied by a strong reduction of the North Atlantic poleward heat transport (at 30°N) from 0.7 PW to 0.1 PW (1PW = 10^{15} W). This leads to a strong surface cooling over the North Atlantic and Europe (Fig. 1) and to increased sea-ice cover. The sea surface temperature (SST) cools by almost 4 K averaged over the entire North Atlantic (Fig. 2b). After the meltwater input was stopped, the convection slowly recovered, and the Atlantic circulation returned to the conveyor-belt type overturning pattern. The North Atlantic SST warms up within a few decades by 2.5 K (Fig. 2b).

Although the meltwater responses have maximal amplitude in the Atlantic and Europe, the signal is distinct in the Pacific as well (Fig. 1). Sea surface temperature, averaged over the entire North Pacific, shows a marked cooling, with maximal amplitude of $\sim 2 \text{ K}$ in the third century of the simulation (Fig. 2b). The temperature changes are not as abrupt as in the Atlantic, but the reinitiation of NADW formation shows up as a warming of 1 K within one century. The strong changes in SST and sea-ice distribution lead also to significant changes in the atmospheric circulation. The first empirical orthogonal function (EOF) of the wind stress fields of the Northern Hemisphere Pacific shows a strong cyclonic signal centred around 60°N and 155°W , with westerlies extending across

the Pacific around 45°N . There is a strong northward component along the Canadian and US west coast and a weaker westward component at the Mexican coast (Fig. 3a). The pattern is consistent with an eastward shift in the position of the Aleutian low-pressure area and the intensification of cyclone activity, especially in the eastern part of the North Pacific. The associated principle component (PC) time series (Fig. 2c) shows marked differences to the control run between years 150 and 530 that are clearly related to changes in the Atlantic overturning circulation.

The divergence of the resulting Ekman transports causes anomalous upwelling in the Pacific north of 50°N during periods of suppressed NADW formation. At the west coast of North America, however, the northward wind-stress anomaly causes a convergence of Ekman transports and thus anomalous downwelling along the coast. The corresponding negative anomaly of the vertical velocities extends down to depths of 850 m. At 1,500 m, however, the anomaly changes sign and shows an upward component in the northeast Pacific. The pattern of the associated response in SST (Fig. 3b) reveals a general cooling of the North Pacific north of 30°N . Maximal cooling ($\sim 2 \text{ K}$) can be seen in a tongue around 50°N extending from 140°E to 150°W . This pattern corresponds with an enhanced cold-air advection from Siberia, especially in the winter season. The relatively small cooling at the American coast coincides with the intensification of the northward-blowing winds along the coast and the reduction of the upwelling of colder subsurface waters south of 40°N .

In an uncoupled sensitivity experiment with the ECHAM3 atmosphere model, the SST and sea ice were prescribed according to the meltwater experiment in the Atlantic and to the control run elsewhere. The atmospheric response over the Atlantic and Eurasia (not

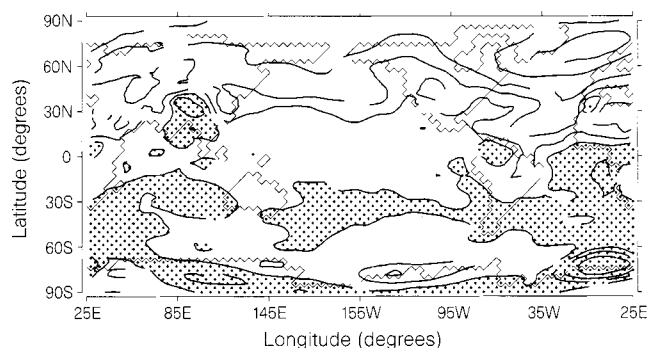


Figure 1 Change in decadal mean near-surface air temperature in the case of a shutdown of North Atlantic Deep Water formation in the ECHAM3/LSG coupled OAGCM. Displayed are the mean differences between two 10-year synchronous integrations corresponding to years 241 to 250 from both meltwater experiment and control run (see text). Details of the experiments are described in ref. 18 and time series of integral quantities are shown in Fig. 2. Isolines are plotted for $\pm 0, 1, 2, 4, 8$ and 16 K . Stippling indicates higher temperatures in the meltwater experiment. Note significant cooling in the North Pacific north of $30\text{--}40^\circ \text{N}$.

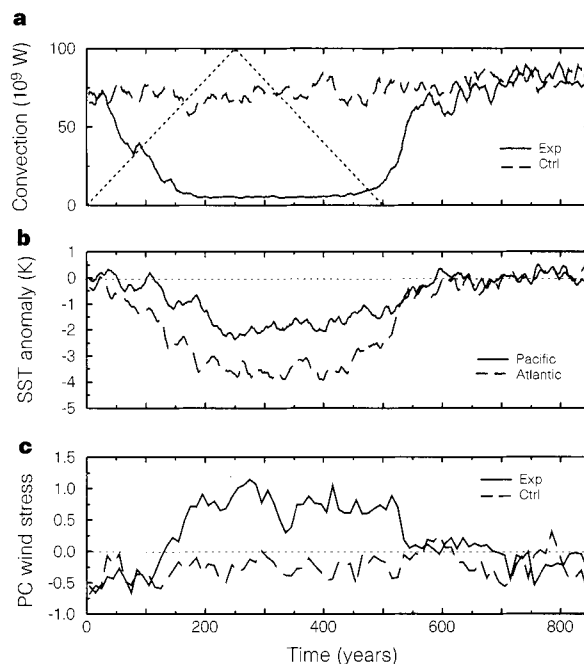


Figure 2 Time series (850 years) from meltwater experiment (Exp) and control run (Ctrl). All data are decadal mean data without further filtering. **a**, Convection in North Atlantic and Arctic (an indicator of the actual NADW formation rate) and a schematic illustration of the prescribed meltwater input (dotted line) with a maximum of 0.625 Sv in year 250. **b**, SST anomalies between meltwater experiment and control. The solid line shows the SST averaged over the Pacific north of 30°N . The dashed line shows the time series averaged in the Atlantic between 30° and 70°N . Locally the annual mean cooling reaches maxima of more than 8 K . **c**, Principal component (PC) time series of the wind stress EOF shown in Fig. 3a.

shown) was very similar to the response in the meltwater experiment (Fig. 1). The atmosphere extracted large amounts of heat near the northwest Pacific coast due to the outflow of colder air from Siberia in winter. During the period of cooling around year 200 in the coupled simulation, the annual mean heat loss from the North Pacific Ocean to the atmosphere increased by 0.2 PW compared to the control experiment. During the rest of the cold period this additional heat loss is reduced to 0.1 PW. Thus the North Pacific cooling is a direct consequence of the climate changes in the North Atlantic. The changes in atmospheric circulation over the North Pacific found in the meltwater experiment is a combination of the far-field response from the Atlantic and the local interaction with the increased meridional SST gradient over the North Pacific.

In the coupled model, the cooling in the surface layer also leads to an intensified ventilation of the thermocline in the North Pacific, as was already found previously in a zonally averaged model²⁰. Especially in the northeast Pacific, the wintertime mixed layer depth is increased. The ventilation of the thermocline produces a tongue of colder and fresher water which penetrates deeper and further to the south than in the control simulation before it leaves the coast flowing westward. In the zonally averaged mass-transport stream function this shows up as an increase of 3 Sv of the North Pacific Intermediate Water formation (not shown). We illustrate the effect of the circulation changes on the ventilation using an offline

advective tracer model for radiocarbon¹⁴. The temporal evolution of the resulting difference in $\Delta^{14}\text{C}$ between the surface and 450 m depth (approximately sill depth in the Santa Barbara basin⁶) along 35° N in the Pacific is shown in Fig. 4. Along the American coast, the reduction of the $\Delta^{14}\text{C}$ differences is almost 30‰. This can be explained by the combined effect of wind-induced reduction of the upwelling of old deep water and the deeper penetration of young water. This signal is consistent with the changes found by Kennett and Ingram⁶, except that our model overestimates the ventilation age. In the model, the radiocarbon signal at the coast at 450 m depth is associated with a cooling and freshening (~ 0.8 K and 0.4 practical salinity units). There is, however, a strong depth dependence of the radiocarbon signal near the coast. At the next model level (250 m depth) the $\Delta^{14}\text{C}$ difference with the surface is reduced by 25‰ during the cold period. At 2,000 m depth, the radiocarbon concentrations indicate an increase in ^{14}C depletion along the American coast, but with a clear delay compared to the surface response. This depth-dependent pattern of ventilation changes appears to be consistent with new data from deeper sites in the northeast Pacific.⁷ However, for comparison with observed data, one also must take into account that the atmospheric radiocarbon concentration increased ~ 35 ‰ during the Younger Dryas²⁷. This response has been successfully simulated with simplified coupled models^{21,28}.

In a sensitivity experiment with the uncoupled ocean model with

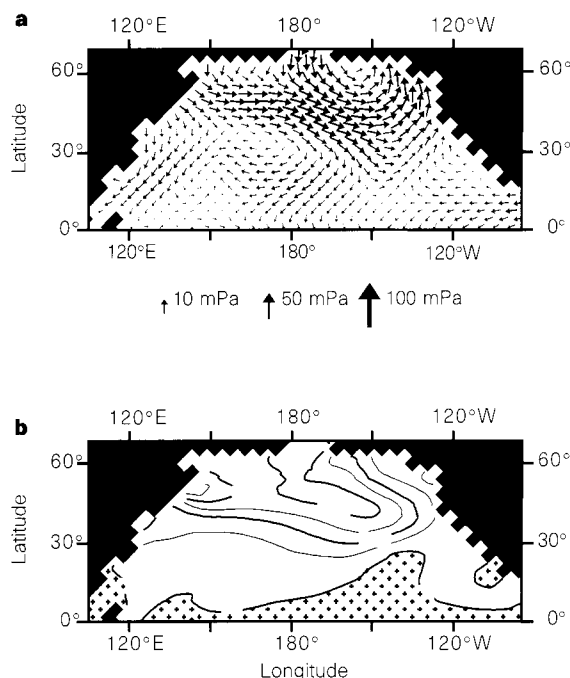


Figure 3 **a**, First common EOF (explains 38.5% of total variance) of Northern Hemisphere Pacific wind stress from both control and meltwater experiment and **b**, associated correlation pattern of SST (in K). Decadal means of the data from both the meltwater experiment and the control run were used for this analysis. The common time-mean value from both experiments was subtracted before the analysis. To obtain the strength of the anomaly for any particular time interval of the simulation, both patterns must be multiplied by the difference of the loading of the principal component (PC) shown in Fig. 2c. The pattern is strongest in the winter season. The change in SST associated with this response was computed by linear regression analysis with the PC of the wind stress EOF. As the wind stress pattern is directly related to the SST in the North Atlantic, this pattern shows the combined effect of the advection of colder air from Siberia (compare with Fig. 1) onto the North Pacific and the temperature effects associated with changes in the atmospheric circulation.

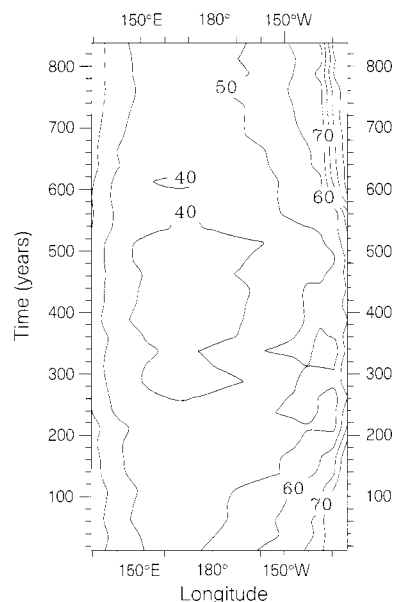


Figure 4 Hovmöller diagram of radiocarbon ($\Delta^{14}\text{C}$) differences between surface and 450 m depth at 35° N in the Pacific from an advective tracer model. ($\Delta^{14}\text{C}$ is defined as the deviation of the $^{14}\text{C}/^{12}\text{C}$ ratio from a given standard and corrected for fractionation effects, given in ‰). Shown are averages over 25 years of the meltwater experiment; contour interval is 10‰. Shading indicates values >60 ‰. The results are obtained using an offline advective tracer model, which was spun-up by making a sequence of twenty 850-year integrations using the flow fields from the individual years of the control run. Each run was started with the $\Delta^{14}\text{C}$ field achieved at the end of the previous simulation. The model was then forced with flow fields from the 850 years of the meltwater experiment. As the change in surface concentrations at this latitude was generally <10 ‰ and showed younger radiocarbon values during the cold period, the reduction in $\Delta^{14}\text{C}$ difference by ~ 10 ‰ over the whole Pacific reflects mainly changes at 450 m depth. Further south (for example, at 28° N) the model shows a similar increase in the ventilation of the central Pacific, but changes in coastal areas are not as large.

a restoring time constant of 5 months for the surface temperature¹⁶, a collapse of NADW formation induced by the introduction of a large negative salinity anomaly in the North Atlantic leads also to enhanced ventilation of the thermocline with fresher water (with signal propagation by coastal and equatorial Kelvin and Rossby waves within less than three decades from the Atlantic through the Indian Ocean to the northeast Pacific). However, this mechanism (compare ref. 6) accounts for only about one-third of the radiocarbon signal in the OAGCM. The remaining part thus can be explained by effects caused by changes in the atmosphere.

Although there is some agreement between model and geological data for circulation changes along the American west coast, there are also some disagreements. The cooling near the coast in our model is rather small (~1 K), whereas sediment records suggest^{6,29} a surface cooling of 2–3 K. Because there is some discrepancy between alkenone-based and foraminifera-based (*Globigerina pachyderma*) SST estimates for the Last Glacial Maximum from the Santa Barbara basin^{6,29}, and *G. pachyderma* may have a subsurface habitat³⁰, the upwelling 'discrepancies' may not necessarily reflect model inadequacies. The large changes in land ice cover, which are not included in our simulations, would also influence the circulation response. Another important shortcoming of the simulations is the coarse resolution (both horizontally and vertically) which makes the comparison with local phenomena in regions with large topography gradients questionable and does not allow the simulation of several important small-scale features. To address the above discrepancies a more complete set of experiments would be required with fully realistic boundary conditions for the deglacial.

Despite the differences mentioned above, our results clearly demonstrate the influence of variations in NADW formation on the North Pacific. In the case of a collapse of NADW formation, both the atmospheric and oceanic transmission of the signal lead to enhanced ventilation of the northeast Pacific thermocline, with the atmospheric effect about twice as strong as the oceanic. These results explain concurrent changes in the North Atlantic and North Pacific for both the Younger Dryas and (possibly) earlier millennial-scale cooling events of a similar nature^{9–12}. Owing to the atmospheric teleconnection, a cooling in the North Atlantic and increased sea-ice cover alone seem to be sufficient to enhance thermocline variation in the northeast Pacific. □

Received 30 September 1996; accepted 2 April 1997.

1. Hessler, C. J. *Late Pleistocene Environments of North Pacific and North America* (Spec. Publ. 35, Geogr. Soc., New York, 1960).
2. Petet, D. M. & Mann, D. H. Late-glacial vegetation, tephra, and climatic history of southwestern Kodiak Island, Alaska. *Ecoscience* **1**, 255–267 (1994).
3. Petet, D. M. Global Younger Dryas? *Quat. Int.* **38**, 93–104 (1995).
4. Kalle, N. *et al.* Evidence of cooling during the Younger Dryas in the western North Pacific. *Oceanol. Acta* **11**, 369–375 (1988).
5. Keigwin, L. D. & Jones, G. A. Deglacial climatic oscillations in the Gulf of California. *Paleoceanography* **5**, 1009–1023 (1990).
6. Kennett, J. P. & Ingram, B. L. A 20,000-year record of ocean circulation and climate change from the Santa Barbara Basin. *Nature* **377**, 510–514 (1995).
7. van Geen, A. *et al.* Ventilation changes in the northeast Pacific during the last deglaciation. *Paleoceanography* **11**, 519–528 (1996).
8. Thunell, R. C. & Mortyn, P. G. Glacial climate instability in the Northeast Pacific Ocean. *Nature* **376**, 504–506 (1995).
9. Behl, R. J. & Kennett, J. P. Brief interstadial events in the Santa Barbara Basin, NE Pacific, during the past 60 kyr. *Nature* **379**, 243–246 (1996).
10. Kotilainen, A. T. & Shackleton, N. J. Rapid climate variability in the North Pacific during the past 95,000 years. *Nature* **377**, 323–326 (1995).
11. Benson, L. V. *et al.* Climatic and hydrologic oscillations in the Owens Lake Basin and adjacent Sierra Nevada, California. *Science* **274**, 746–749 (1996).
12. Philips, F. M. *et al.* Chronology for fluctuations in Late Pleistocene Sierra Nevada glaciers and lakes. *Science* **274**, 749–751 (1996).
13. Rooth, C. Hydrology and ocean circulation. *Prog. Oceanogr.* **11**, 131–149 (1982).
14. Maier-Reimer, E. & Mikolajewicz, U. in *Oceanography 1988* (eds Ayala-Castañares, A., Wooster, W. & Yáñez-Arancibia, A.) 87–100 (UNAM, Mexico, 1989).
15. Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
16. Mikolajewicz, U. & Maier-Reimer, E. Mixed boundary conditions in ocean general circulation models and their influence on the stability of the model's conveyor-belt. *J. Geophys. Res.* **99**, 22633–22644 (1994).
17. Manabe, S. & Stouffer, R. J. Simulation of abrupt climate change induced by freshwater input to the North Atlantic Ocean. *Nature* **378**, 165–167 (1995).
18. Schiller, A., Mikolajewicz, U. & Voss, R. The stability of the North Atlantic thermohaline circulation in a coupled ocean-atmosphere general circulation model. *Clim. Dyn.* (in the press).

19. Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 941–966 (1988).
20. Wright, D. G. & Stocker, T. F. *Ice in the Climate System* (ed. Peltier, W. R.) 395–416 (Springer, Berlin, 1993).
21. Stocker, T. F. & Wright, D. G. Rapid changes in ocean circulation and atmospheric radiocarbon. *Paleoceanography* **11**, 773–795 (1996).
22. Roeckner, E. *et al.* Simulation of the present-day climate with the ECHAM model: impact of model physics and resolution. (Tech. Rep. No 93, Max-Planck-Inst. für Meteorologie, Hamburg, 1992).
23. Maier-Reimer, E., Mikolajewicz, U. & Hasselmann, K. Mean circulation of the LSG OGCM and its sensitivity to the thermohaline surface forcing. *J. Phys. Oceanogr.* **23**, 731–757 (1993).
24. Voss, R. & Sausen, R. Techniques for asynchronous and periodically synchronous coupling of atmosphere and ocean models. Part II: impact of variability. *Clim. Dyn.* **12**, 605–614 (1996).
25. Voss, R., Sausen, R. & Cubasch, U. Periodically synchronously coupled integrations with the atmosphere-ocean general circulation model ECHAM3/LSG. *Clim. Dyn.* (submitted).
26. Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
27. Goslar, T. *et al.* High concentration of atmospheric ¹⁴C during the Younger Dryas cold episode. *Nature* **377**, 414–417 (1995).
28. Mikolajewicz, U. A meltwater induced collapse of the 'conveyor belt' thermohaline circulation and its influence on the distribution of ¹⁴C and ¹⁸O in the oceans. (Tech. Rep. 189, Max-Planck-Inst. für Meteorologie, Hamburg, 1996).
29. Herbert, T. D., Yasuda, M. & Burnett, C. Glacial-interglacial sea-surface temperature record inferred from alkenone unsaturation indices, site 893, Santa Barbara Basin. *Proc. ODP* **146** (pt 2), 257–264 (1995).
30. Sautter, L. R. & Thunell, R. C. Seasonal variability in the ¹⁸O and ¹³C of planktonic foraminifera from an upwelling environment: Sediment trap results from the San Pedro Basin, Southern California Bight. *Paleoceanography* **6**, 307–334 (1991).

Acknowledgements. We thank T. Herbert, L. Keigwin, E. Maier-Reimer, T. Stocker and L. Stott for discussions. T.J.C. was supported by the US NSF.

Correspondence should be addressed to U.M. (e-mail: mikolajewicz@dkrz.de).

Catastrophic collapse at stratovolcanoes induced by gradual volcano spreading

B. van Wyk de Vries & P. W. Francis

Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes, UK

Unlike ordinary mountains, which are formed by slow uplift and erosion, volcanoes are constructed rapidly. As a consequence, many are liable to massive flank failures, leading to debris avalanches (for example, at Mount St Helens in 1980). Such failures occur worldwide about once every 25 years (ref. 1) and even small ones can present a major hazard—in particular if far-reaching tsunamis are generated, as at Mayu-yama in 1792 (ref. 2). Previous work has tended to emphasize differences in eruption style associated with flank failure², but here we focus on the fundamental structural causes of failure. Most volcanic failures are generated by magmatic intrusion and flank spreading³. We present evidence, however, that Mombacho volcano in Nicaragua experienced a previously unrecognized type of failure, triggered by sub-volcanic basement spreading. Notably, collapses related to basement spreading do not require that the volcano be magmatically active, and thus flank failure may pose a significant risk even in inactive volcanoes, which are rarely monitored.

Mombacho volcano rises 1,400 m above the west shores of Lake Nicaragua, on a basement of Quaternary ignimbrite of the Las Sierras Formation^{4,5} (Fig. 1). Conspicuous debris avalanche deposits on two sides of the volcano provide unequivocal evidence for recent flank failure. One deposit, below a well defined collapse scar, forms the remarkable Las Isletas archipelago in Lake Nicaragua (Fig. 1). The other, on the south side below the deep 'El Crater' scar, covers ~60 km² with hummocky avalanche deposit (Fig. 2). Forest cover indicates that the avalanche deposits are at least 1,000 years old. The 20,000-yr-old Apoyo pumice underlies the deposits, giving a maximum age⁶.

Volcano spreading provides a framework to interpret the structural and magmatic evolution of volcanoes^{7–9}. We propose here a new failure mode caused by lateral displacement of sub-volcanic strata. A similar process has been proposed for the Hawaiian shield

volcanoes¹⁰ but remains controversial^{11,12}, in part because of the lack of physical evidence for the involvement of basement material. Our observations lend weight to the hypothesis, although we stress the differences in scale and geological context: Mombacho is a small standard stratocone, whereas Hawaii is a large oceanic shield complex. In particular, the thickness of the proposed weak layer at Hawaii is a small fraction (1%) of the volcano height⁷, and acts purely as a slip plane on which internal deformation of the volcano flank is accommodated. At Mombacho, the weak layer thickness is ~20% of the volcano height and it is deformation taking place *within* this layer that causes failure of the overlying edifice.

Earlier work suggested that the radial basement spreading limited the hazard of flank failure at volcanoes⁴, but here we argue that spreading on a restricted sector initiated the Las Isletas failure. In contrast, the El Crater failure resulted from progressive hydrothermal weakening of the cone and consequential flank spreading. Evidence for explosive activity in the form of tephra layers, pyroclastic flow deposits or juvenile fragments is absent from both the Las Isletas and El Crater deposits. This emphasises that flank failure can occur without associated magmatic activity.

The presence of large amounts of Las Sierras ignimbrite within the Las Isletas avalanche deposit demonstrates that the failure involved basement, because these ignimbrites underlie the volcano. The Las Sierras material comprises the central and lowermost parts of the deposit, adjacent to a shallow basin forming the Aseses inlet (Figs 1 and 2). A small scarp extending from the eastern wall to the lake indicates that the failure margin reached the lake here. This suggests that the failure plane cut into basement under Aseses beyond the volcano foot. An analogous situation occurs at Socompa, North Chile, where a debris avalanche deposit contains 60% basement¹³ and the failure margin extends 5 km from the volcano^{13,14}. Blocks of Las Sierras ignimbrite have pervasive calcite-cemented shear fractures, quite distinct from the open, spaced fractures associated with avalanche disaggregation, and consistent with deformation before failure. Thrusting along the north base and slumping of the southeastern side⁴ provide additional evidence of slow basement deformation (Fig. 2). Thus before collapse took place, the north and east base of Mombacho was spreading outwards. This preferential direction of spreading may reflect localized lake sediment layers within the Las Sierras Formation, though these were not identified in the deposit. Alternatively, the orientation of spreading structures could have been controlled by regional stress, which in Nicaragua promotes east–west extension and north–south compression¹⁵.

The floor of the collapse scar slopes at a similar angle to the cone surface, suggesting that the avalanche slid along a bedding plane (Fig. 3). In this respect, Las Isletas resembles many dip-slope slides and avalanches, both volcanic and non-volcanic^{16,17}, but it also has a lower part extending into the basement. The combination of dip-slope upper slide and a lower spreading sector could have developed into the avalanche in the way shown in Fig. 3a. Thrust anticline growth at the front of the spreading sector elevates the front of the spreading sector, so that basement can be included in a subsequent failure without excessively deep excavation (Fig. 3a).

The model implies that spreading is integral to the development of failure and demonstrates that identification of spreading is an important first step in assessing potential collapse hazard. Furthermore, in contrast to radially spreading volcanoes⁴, preferential spreading in one direction is critical to collapse development: whereas radial spreading tends to generate inward-dipping graben which inhibit collapse⁸, sector spreading (as at Mombacho) generates failure-prone outward-dipping structures (Fig. 3b). Spreading in a preferential direction may be caused by buttressing⁸, by the regional slope of basement beds, by regional stress¹⁵, by weak basement or by high fluid pressures under one side. Importantly, hot springs in the Aseses inlet suggest a possible role for elevated pore pressures (Fig. 2).



Figure 1 View of the Las Isletas collapse deposit at Mombacho. The islands of the archipelago are the distal part of the avalanche deposit and are composed of basement Las Sierras ignimbrite with a veneer of rocks from the volcano above. The basement ignimbrites were excavated from the Aseses embayment seen in the middle ground. They slope from 150 m above sea level on the western side of Mombacho to about sea level in the area of the photograph: a dip of ~2° (ref. 4). There are no confirmed reports of historic eruptions at Mombacho²⁵.

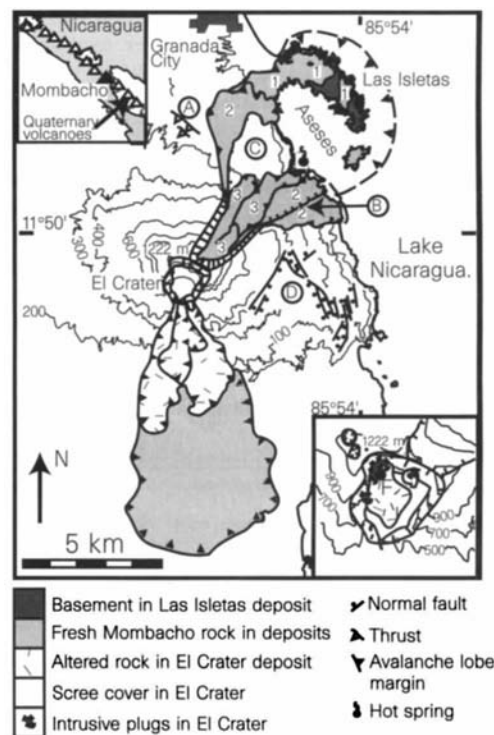


Figure 2 Geological map of Mombacho, showing the two recent flank failures and debris avalanche deposits of Las Isletas and El Crater. Insets show location of Mombacho and a detail of El Crater. Points on map draw attention to: A, thrusting observed at north base⁴; B, scarp in Las Isletas deposit indicating buried failure margin extending down into basement; C, La Calera post-collapse lava flow; D, normal faults resulting from spreading of the southeastern flank. Fumarole field in El Crater shown by black circle at F. Numbers in Las Isletas deposit indicate sequential emplacement of lobes by retrogressive collapse. Within the central lobe which forms the Las Isletas archipelago (1), the Las Sierras facies forms the lowest material, partially covered by fresh lava blocks. By contrast, the overlying marginal lobes are composed entirely of fresh cone material (2). Small secondary lobes overlie the La Calera lava flow, erupted in the scar after the collapse (3), and one large headwall segment still awaits collapse.

In contrast to Las Isletas, the El Crater collapse scar is at least 400 m above the sub-volcanic basement, so failure clearly took place wholly within the cone. The proximal avalanche deposit is composed of thick lobes of a chaotic mixture of hydrothermal clays, altered lava and scoria. Buttresses on the walls of the scar are plugs of basalt and andesite, in places highly altered. Below one of these there is a vigorous fumarole field. A zone of intense alteration extends over much of the lower crater, the material being similar to that in the avalanche. Progressive alteration of the volcano core, converting

strong volcanic and intrusive rock to weak hydrothermal clays¹⁸, probably caused collapse (Fig. 3c)¹⁹.

Because hydrothermal weakening is progressive, deformation of affected volcanoes is likely to begin long before failure, as in other deep-seated failures¹⁷, and would be indicated by steepening and thrusting of the lower flank and by normal faulting higher up. Increased fluid pressure²⁰ and magma intrusion could accelerate failure, with pre-existing mass creep controlling the failure direction. Swanson²¹ suggested such a relationship between old altered domes, and the orientation of the Mt St Helens collapse.

Precursory slow spreading probably triggered flank failures at other volcanoes, notably Socompa¹⁴. Probable precursory spreading has already been detected at Colima²² and Etna²³, and we postulate future failures for other volcanoes currently exhibiting evidence of active spreading (Fig. 4). Spreading features observed at Kilauea may be sites of potential collapse, though the direct involvement of the basal décollement in such events is still uncertain¹⁰⁻¹².

Because the features described here are produced by gradual processes, they may be detectable long before any failure occurs. Identification of basement and flank spreading allows prediction of the type and location of potential collapses and assessment of their probable effects. Potential collapses can be detected from evidence of individual spreading sectors (Fig. 4a), while the symptoms of flank spreading will be bulging and fracturing (Fig. 4b), core weakening by persistent fumarolic activity, and evidence of old domes and plugs. At Mt St Helens, where collapse was immediately preceded by considerable deformation²⁴, evidence of previous flank movement in the collapse scar²¹ suggests that if the north flank had been monitored before 1980, the location of the ultimate collapse could have been identified well before the recent activity.

Once the location and structure of potential collapses are recognized, deformation monitoring can determine baseline rates of spreading activity, which probably range from 1 to 50 cm yr⁻¹

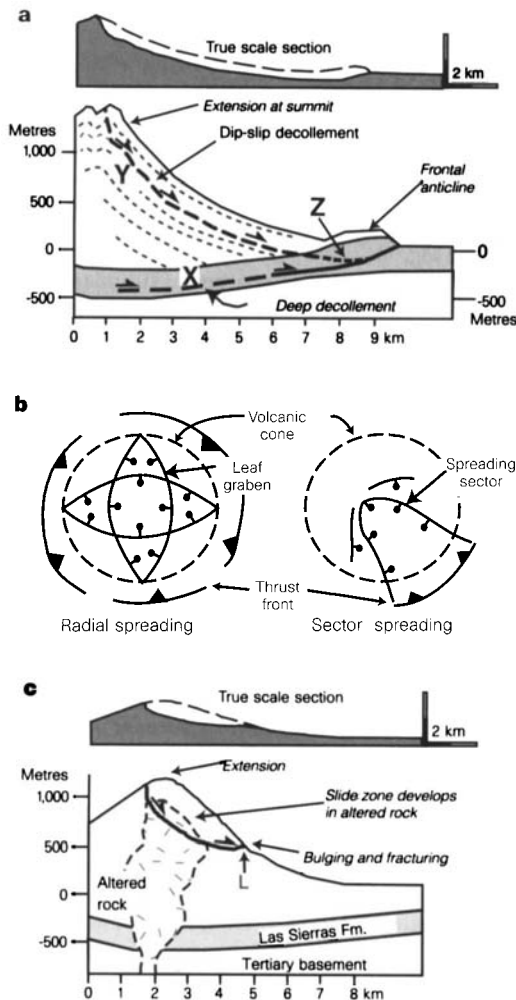


Figure 3 Interpretation of spreading and collapse structures at Mombacho, showing contrasting models for basement failure at Las Isletas (**a** and **b**), El Crater (**c**) and probable precursory features. **a**, Las Isletas: basement collapse. Deep spreading on decollement (X) within the Las Sierras Formation rises to produce a frontal anticline at the foot of Mombacho. Spreading induces differential movements within the upper parts of the volcano⁴, contributing to the initiation of slip on a plane (Y) within the cone. Once the dip-slip decollement is activated, the mass above it places additional load on the spreading front, inducing increased movement, and eventual failure through the frontal anticline (Z). **b**, Plan diagrams illustrating the difference in structural style between a radially spreading and a sector spreading volcano (that is, Mombacho). Radial spreading produces inward dipping normal faults that cut any potential failure plane in the cone^{4,8}. In contrast, sector spreading creates outward dipping faults, which promote collapse. **c**, El Crater: flank collapse. The El Crater collapse crater is ~1.5 km wide and long, and 700 m deep. Its walls curve inwards toward the opening, where there is a pronounced 30-m-high lip (L). Overall, the shape is that of a rotational slump failure in a mechanically homogenous medium¹⁷, so no pre-existing decollement plane is required. Such failures usually begin when shear strength is reduced over a wide area^{17,26}. At El Crater this homogeneous strength loss was produced by progressive hydrothermal alteration.

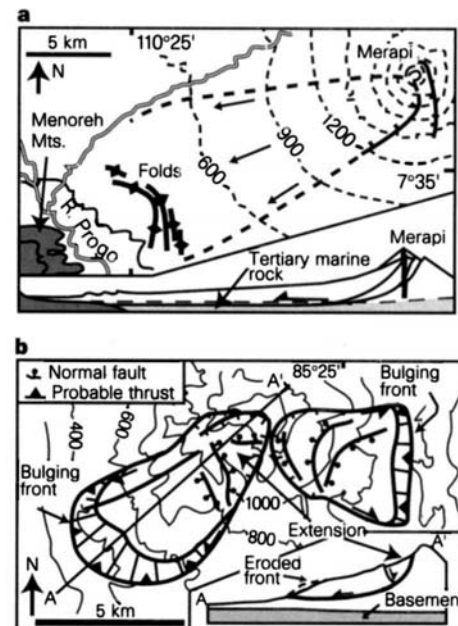


Figure 4 Examples of potential failures at spreading volcanoes. **a**, A spreading sector with collapse potential: Merapi, Java; a long sector is spreading. Drawn after van Bemmelen⁹. The compressive front is restrained by the Menoreh mountains. This barrier may have inhibited flank failure during a rapid spreading stage before a major eruption⁹. **b**, A spreading flank with collapse potential: Orosi, Costa Rica; two flank sectors are creeping outwards. Each has a steepened front just above the cone foot, probably hosting a thrust, and a crescent-shaped graben higher up.

(refs 4, 9), and unusual accelerations which may immediately precede collapse can be detected. Hundreds of volcanoes are potentially liable to collapse: prediction of future collapse sites depends on their timely identification by the criteria presented here, and subsequent monitoring of deformation. As population densities increase on the flanks of volcanoes around the world, recognition and prediction of flank failure becomes increasingly urgent. □

Received 3 October 1996; accepted 15 April 1997.

1. Siebert, L. Threats from debris avalanches. *Nature* **356**, 658–659 (1992).
2. Siebert, L., Glicken, H. & Ui, T. Volcanic hazards from Bezymianny- and Bandai-type eruptions. *Bull. Volcanol.* **49**, 435–459 (1987).
3. Denlinger, R. P. & Okubo, P. Structure of the mobile south flank of Kilauea Volcano, Hawaii. *J. Geophys. Res.* **100**, 24499–24507 (1995).
4. van Wyk de Vries, B. & Borgia, A. in *Volcano Instability on the Earth and Other Planets* (eds McGuire, W. J. et al.) 95–110 (Spec. Publ. 110, Geol. Soc., London, 1996).
5. Hradecky, P. *Geología del Volcán Mombacho* (INETER archives, Managua, Nicaragua, 1987).
6. Sussman, D. Apoyo Caldera, Nicaragua: a major Quaternary silicic eruptive center. *J. Volcanol. Geotherm. Res.* **24**, 249–282 (1980).
7. Borgia, A. The dynamic basis of volcanic spreading. *J. Geophys. Res.* **99**, 17791–17804 (1994).
8. Merle, O. & Borgia, A. Scaled experiments of volcanic spreading. *J. Geophys. Res.* **101**, 13805–13817 (1996).
9. van Bemmelen, R. W. *Mountain Building* (Martinus Nijhoff, The Hague, 1954).
10. Iverson, R. M. Can magma-injection and groundwater forces cause massive landslides on Hawaiian volcanoes? *J. Volcanol. Geotherm. Res.* **66**, 295–308 (1985).
11. Elsworth, D. & Voight, B. in *Volcano Instability on the Earth and Other Planets* (eds McGuire, W. J. et al.) 45–53 (Spec. Publ. 110, Geol. Soc. London, 1996).
12. Moore, J. G. et al. Prodigious submarine landslides on the Hawaiian Ridge. *J. Geophys. Res.* **94**, 17465–17484 (1989).
13. Wadge, G., Francis, P. W. & Ramirez, C. F. The Socompa collapse and avalanche event. *J. Volcanol. Geotherm. Res.* **66**, 309–336 (1995).
14. van Wyk de Vries, B., Self, S. & Francis, P. W. Was Socompa pushed or did it fail? *J. Conf. Abstr.* **2**(1), 78 (1997).
15. van Wyk de Vries, B. & Merle, O. The effect of volcanic constructs on rift fault patterns. *Geology* **24**, 643–646 (1996).
16. Naranjo, J. A. & Francis, P. W. High velocity debris avalanche deposit at Lastarria volcano in the north Chilean Andes. *Bull. Volcanol.* **49**, 509–514 (1987).
17. Chigira, M. Long-term gravitational deformation of rocks by mass rock creep. *Eng. Geol.* **32**, 157–184 (1992).
18. Lopez, D. L. & Williams, S. N. Catastrophic volcano collapse: relation to hydrothermal alteration. *Science* **260**, 1794–1796 (1993).
19. Siebert, L. Large volcanic debris avalanches: characteristics of source areas, deposits, and associated eruptions. *J. Volcanol. Geotherm. Res.* **22**, 163–197 (1984).
20. Reid, M. E. Destabilizing hydrothermal pressurisation in volcanoes. *Abstr. IUGG XXI Assembly Vol. A*, 464 (Int. Un. Geodesy Geophys., Boulder, 1995).
21. Swanson, D., Hausback, B. P. & Zimbelman, D. R. Why was the 1980 Bulge on the North flank of Mount St. Helens? *Abstr. IUGG XXI Assembly Vol. A*, 464 (Int. Un. Geodesy Geophys., Boulder, 1995).
22. Murray, J. B. Ground deformation at Colima volcano 1982–1991. *Geophys. Int.* **32**, 659–670 (1994).
23. Murray, J. B., Voight, B. & Glot, J.-P. Slope movement crisis on the east flank of Mt. Etna volcano: Models for eruption triggering and forecasting. *Eng. Geol.* **38**, 185– (1992).
24. Moore, J. G. & Albee, W. C. *Prof. Pap. US Geol. Surv.* **1250**, 123–141 (1981).
25. Simkin, T. & Siebert, L. *Volcanoes of the World* (Geoscience, Tuscon, 1995).
26. Terzaghi, K. & Peck, R. B. *Soil Mechanics in Engineering Practice* 2nd edn (Wiley, New York, 1967).

Acknowledgements. We thank W. Strauch, P. Perez and O. Membréniz for assistance in the field, and H. Rymer, D. Rothery, D. Peate, F. Garland and P. Bland for reviews. This work was supported by the Leverhulme Trust.

Correspondence and requests for materials should be addressed to B.v.W. d.v. (e-mail: b.van-wyk-de-vries@open.ac.uk).

New evidence concerning avian origins from the Late Cretaceous of Patagonia

Fernando E. Novas* & Pablo F. Puerta†

* Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Av. Angel Gallardo 470, Buenos Aires (1405), Argentina

† Museo Paleontológico Egidio Feruglio, Av. 9 de Julio 655, (9100) Trelew, Argentina

The spate of recent discoveries of Mesozoic birds has substantially improved our understanding of the early evolution of birds and flight^{1–5}, but has failed to close the morphological gap between the Upper Jurassic *Archaeopteryx lithographica*, the earliest known bird, and the Dromaeosauridae, the group of non-avian theropod dinosaurs regarded as most closely related to birds^{6,7}. Here we describe a theropod dinosaur from Patagonia, *Unenlagia comahuensis* gen. et sp. nov., which partially fills this gap. Despite

the relatively late appearance of this dinosaur in the fossil record (Upper Cretaceous), several features of *Unenlagia* are more bird-like than in any other non-avian theropod so far discovered. *Unenlagia* resembles *Archaeopteryx* in the morphology of the scapula, pelvis and hindlimb. But several shared, primitive features of the pubis, ischium and hindlimb proportions suggest that *Unenlagia* may represent the sister taxon of the Avialae (=Aves). The structure of the forelimb suggests that the avian mode of forelimb folding, and the extensive forelimb elevation necessary for powered, flapping flight, was already present in cursorial, non-flying theropod dinosaurs.

Theropoda
Coelurosauria⁷
Maniraptora⁷

Unenlagia comahuensis gen. et sp. nov.

Etymology. *Unenlagia*, Latinized from “uñen” and “lag”, Mapuche Indian names respectively meaning “half” and “bird”⁸; and *comahuensis*, from Comahue, a Mapuche name referring to North-West Patagonia.

Holotype. (See Fig. 1).

Locality and horizon. Upper Cretaceous (Turonian-Coniacian⁹), Río Neuquén Formation, Sierra del Portezuelo, Neuquén Province, Argentina. This formation has yielded remains of the basal bird *Patagonykus puertai*^{10,11}, plus a variety of non-avian theropods¹².

Diagnosis. Possesses tall neural spines in posterior dorsals and anterior sacral vertebrae, being nearly twice the height of the centrum; deep lateral pits in the base of the neural spines of these vertebrae; twisted scapular shaft; inflected dorsal margin of post-acetabular iliac blade (Fig. 2).

Unenlagia is a medium-sized maniraptoran dinosaur, nearly 2 m long. Presacral vertebrae are amphiplatyan and have pleurocoels. Six fused sacra are present, although the ilia extend the length of at least nine vertebrae (that is, six sacra, two dorsals, and one caudal), instead of seven as in *Deinonychus* (Museum of Comparative Zoology, MCZ 4871) and *Archaeopteryx*^{13,14}. Proximal haemal arches are craniocaudally short and dorsoventrally long (Fig. 1), resembling those of the dromaeosaurid *Velociraptor*¹⁵.

The scapula (Fig. 2a, b) is strap-like in dorsal aspect and curved in lateral view, closely resembling that of *Archaeopteryx*^{6,16}. As in the latter, the acromion is triangular in lateral aspect, and projected sharply cranioventrally. In contrast to the situation in non-avian theropods (for example, *Deinonychus*^{16,17}), the humeral articulation of the scapula of *Unenlagia* is laterally oriented as in birds^{6,16,18}.

The humerus, estimated to be 27 cm long, is 71% of the femoral length, and is proportionally shorter than in *Archaeopteryx*^{13,14}. As in other Maniraptora⁷, the internal tuberosity (bicipital crest of modern birds¹⁹) is proximodistally extended.

The pelvic bones are not fused (Fig. 2c). The ilium is extensive cranially, but the postacetabular blade is short. The latter is low and sharp as in *Archaeopteryx* and enantiornithines^{14,20}. The fossa for the m. cuppedicus is well developed, a condition that is widely present among Coelurosauria, including basal birds^{10,17,18}. In contrast to non-avian coelurosaurs (that is, *Deinonychus*^{17,21}), the acetabulum of *Unenlagia* tends to close off medially, resembling *Archaeopteryx*¹⁸, *Patagopteryx*²² and hesperornithiforms^{23,24}. The brevis fossa is considerably more reduced than in *Deinonychus*¹⁰. The pubis is slightly shorter than the femur, is oriented ventrally as in other maniraptorans^{10,14}, and distally bears a caudally projected ‘foot’. The pubic shaft expands transversally into a ‘pubic apron’ (Fig. 2d) that is widely extended proximodistally as in *Deinonychus* (MCZ 4371), but in contrast with the more reduced symphysis of *Archaeopteryx* and more derived birds^{6,11}. The ischium is a short, plate-like bone, triangular in side view. The obturator notch is enclosed distally by a triangular obturator process, a primitive condition lost in *Archaeopteryx*^{6,14} and more derived birds^{11,22}. However, the dorsal edge of the ischium exhibits a prominent proximodorsal process (Fig. 2c), separated from the ischiadic

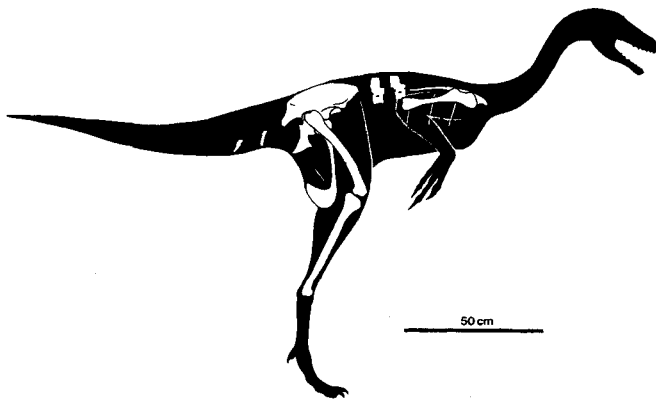


Figure 1 Skeletal reconstruction of *Unenlagia comahuensis*, gen. et sp. nov., based on holotype specimen (MCF PVPH 78; Museo Carmen Funes, Plaza Huincul, Argentina). The specimen was found preserved in partial articulation, with the bones occupying their respective anatomical position. MCF PVPH 78 includes: three presacral vertebrae (presumably correspond to 8th, 10th, and 13th dorsals, the latter articulating to the sacrum), sacrum, dorsal ribs and two proximal haemal arches, left scapula and incomplete humerus, ilia, pubes, and right ischium, right femur and left tibia.

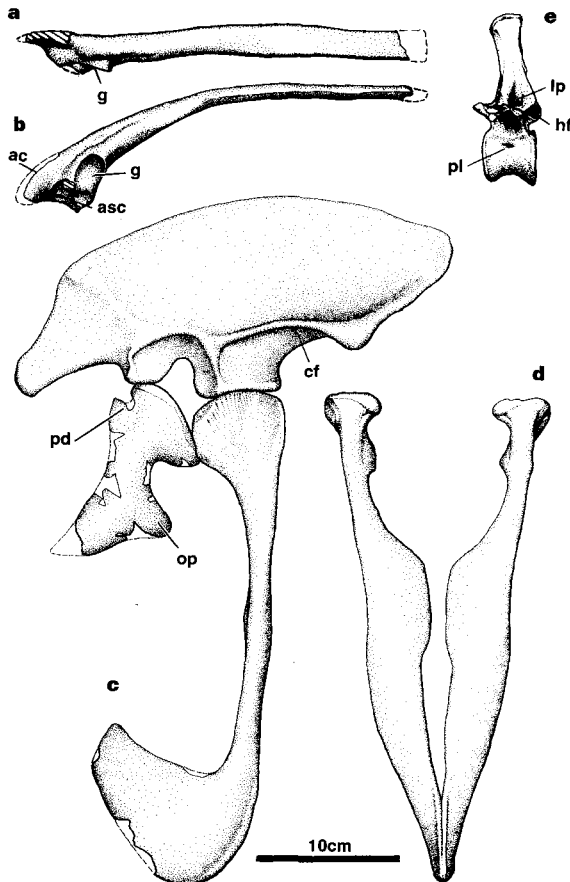


Figure 2 *Unenlagia comahuensis*, gen. et sp. nov. **a, b**, Left scapula in dorsal (**a**) and lateral (**b**) views. **c**, Pelvis in right lateral view. **d**, Pubes in anterior view. **e**, Dorsal 13th in left lateral aspect. Orientation of the scapula shown in **a** and **b** is primarily based on the assumption that the external surface of the acromium is cranio-laterally faced as in other theropods, including birds; thus, the external surface of the scapular blade orientates dorsally (that is, scapular shaft twisted), and the glenoid cavity results laterally oriented. Abbreviations: ac, acromium; asc, articular surface for the coracoid; cf, fossa for origin of m. cuppedicus; g, glenoid cavity; hf, hyposphene; lp, lateral pit; op, obturator process; pd, proximodorsal process on ischium; pl, pleurocoel.

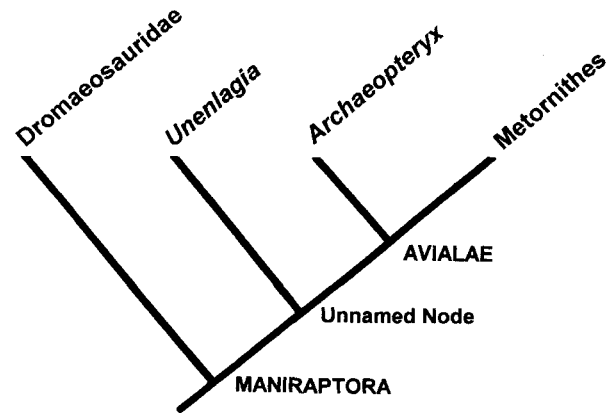


Figure 3 Cladogram depicting phylogenetic relationships of *Unenlagia comahuensis*, gen. et sp. nov. For details, see Methods.

antitrochanter by a deep notch. This character, which is absent in Dromaeosauridae and the remaining non-avian theropods, is similar to that present in *Archaeopteryx*^{3,14,18}.

The femur is long and slender, with a proportionally small head. The anterior trochanter projects proximally and is separated from the greater trochanter by a short cleft, as in *Archaeopteryx*⁶. Both the fourth trochanter and the posterior trochanter^{6,21} are absent. The tibia is long and slender, and is 13% longer than the femur.

Deinonychus was described nearly 25 years ago¹⁷ and its morphology was interpreted as closely resembling that of *Archaeopteryx*^{6,7,25}. No other dinosaur has been discovered until now that closes the morphological gap between these taxa. *Unenlagia* represents such an 'intermediate' dinosaur, because it is more closely related to birds than are dromaeosaurids (Fig. 3). In particular, *Unenlagia* sheds light on the morphology of 'proavian' dinosaurs, clarifying interesting aspects of forelimb function that help us to understand how avian flight emerged.

Some features characterizing the avian mode of forelimb folding were already acquired by non-avian theropods. For example, the pulley-shaped distal carpal present in basal maniraptorans allowed the manus to be flexed against the forearm^{7,17,26}. However, the posteroventrally oriented glenoid cavity of *Deinonychus*¹⁶ (Fig. 4a) constrained the range of humeral movements as much as in other non-avian theropods, preventing it from folding in the avian manner (that is, with the flexor surface of humerus laterally oriented at rest) (Fig. 4b). In *Unenlagia*, instead, the glenoid cavity is laterally oriented (Fig. 4a), a derived feature previously unreported in non-avian dinosaurs¹⁸. This orientation of the glenoid cavity indicates that their forelimbs were capable of excursions in the avian manner (that is, forelimb folding involving rotation of humerus, thus exposing its flexor surface laterally; also, increased humeral abduction allowing an almost vertical positioning of the forelimb during maximum upstroke; Fig. 4b). Interestingly, the lateral orientation of the glenoid cavity of *Archaeopteryx* (Fig. 4a) suggests that this basal bird was also capable of extensive wing elevation (upstroke), as well as fixing the humerus against the body when the wing was held folded. Moreover, such postural activity was critical in protecting the wing feathers.

The large size of *Unenlagia*, together with its proportionally short forelimbs (based on humeral length), argue that this animal was flightless, and its phylogenetic position suggests that it was not derived from a flying form (Fig. 3). Yet *Unenlagia* had already acquired novel, almost avian, forelimb movements, which are lacking in more remote outgroups. The extensive forelimb elevation (upstroke) present in *Unenlagia* constituted a prerequisite for powered, flapping flight, offering the necessary thrust to lift winged theropods from the ground. The design and function of

the forelimbs of *Unenlagia* do not suggest that there was a gliding phase before the acquisition of the avian flight; rather, they are more consistent with the idea that there was a flapping stage^{26,27}, which would require a much greater arc of movement. Whether *Unenlagia* was feathered or not is unknown. The new evidence, however, reinforces the hypothesis that a bipedal, cursorial theropod was ancestral to birds^{6,7}, and used its arms not only in predation, but probably also in maintenance of balance and to control its body attitude while running and leaping^{27,28}.

Remarkably, main differences between forelimb bones of *Archaeopteryx* and those of *Unenlagia*, the closest outgroup of birds, have more to do with relative proportions among bones rather than with the acquisition of structural novelties. In consequence, the development of flying capabilities can be regarded more as changes in body proportions (that is, enlarged forelimbs, reduction in body size and mass) and integument transformations (enlargement of feathers to increase wing surface) than as the acquisition of specific osteological and muscular features. □

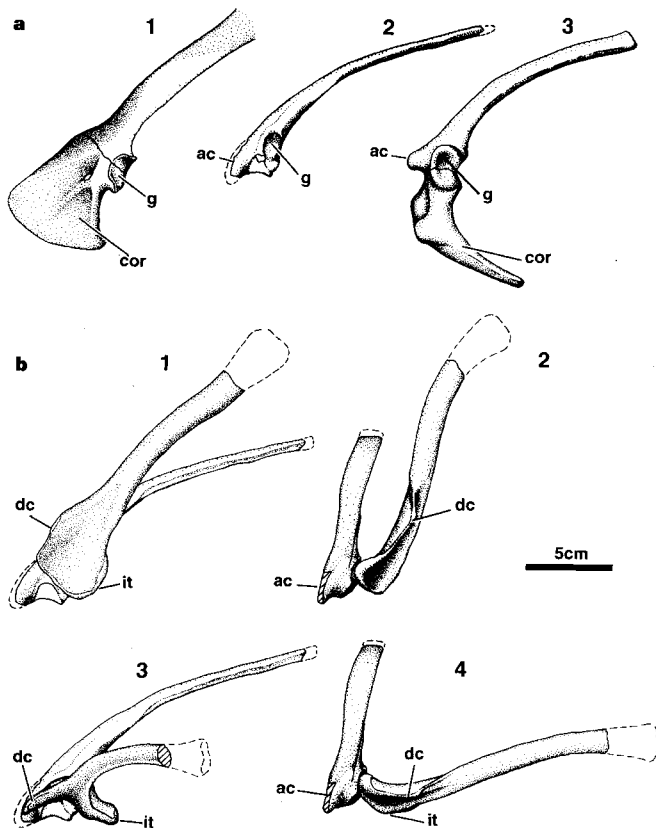


Figure 4 **a**, Lateral view of the left scapula of *Unenlagia* (2), compared with those of *Deinonychus* (1) and *Archaeopteryx* (3) (not to scale; modified from refs 6 and 16). In *Deinonychus*, maximum humeral elevation (=upstroke) was restricted by the posteroventrally oriented glenoid cavity. This is not the case for *Unenlagia* and avialans, in which the glenoid cavity is laterally projected, indicative of extended elevatory range of humeral movements. **b**, Left humerus of *Unenlagia* depicting its maximum elevation (1,2) and depression (3,4) in lateral (1,3) and anterior (2,4) views. Forelimb upstroke was considerably increased dorsally, allowing an almost vertical positioning of the forelimb at maximum abduction. However, the reorientation of the glenoid cavity described for *Unenlagia* and *Archaeopteryx* also implied that maximum humeral depression (=downstroke) was reached when the humerus adopted a nearly horizontal and laterally projected position (3,4). In sum, the outward facing of the glenoid cavity was accompanied by the folding of the forelimbs in a transverse plane, a feature uniquely present in *Unenlagia* and birds among dinosaurs. Abbreviations: ac, acromium; dc, deltopectoral crest; g, glenoid cavity; it, internal tuberosity (=bicipital tubercle).

Methods

The preserved character states of *Unenlagia* were scored into the data matrix (appendix of ref. 11), plus the following characters: acromium triangular and anteroventrally projected; humeral facet laterally oriented; femur elongate, with reduced femoral head; ischium with proximodorsal process on posterior edge; and humerus longer than femur. The resulting 80-character data set was processed using the implicit enumeration option in Hennig 86. The single most parsimonious tree consists of 116 steps, a consistency index of 0.68, and retention index of 0.71. Maniraptoran synapomorphies present in *Unenlagia* are: humerus with proximodistally elongate internal tuberosity, postacetabular blade acuminate, pubis ventrally directed, ischium nearly 50% of pubic length, adductor fossa and associated anterodistal crista of distal femur reduced or absent. No derived characters uniquely shared by *Unenlagia* and Dromaeosauridae were identified. On the contrary, *Unenlagia* shares with Avialae the following unequivocal synapomorphies: femoral fourth trochanter absent; pubic distal foot lacking cranial projection; acromium triangular and projected ventrally; humeral facet of scapula laterally oriented; elongate femur, with reduced femoral head. *Unenlagia* is less derived than *Archaeopteryx* and the remaining avialans because it lacks the following unambiguous characters of Avialae: pubic apron strongly reduced transversely and restricted to the distal 1/3 of the pubic length; obturator process on ischium absent (that is, anteroventral margin of ischium is almost straight); and tibia 25% longer than the femur.

Received 23 December 1996; accepted 27 February 1997.

- Sanz, J. L., Bonaparte, J. F. & Lacasa Ruiz, A. Unusual Early Cretaceous bird from Spain. *Nature* **331**, 433–435 (1988).
- Sereno, P. & Rao, C. Early evolution of avian flight and perching: new evidence from Lower Cretaceous of China. *Science* **255**, 845–848 (1992).
- Chiappe, L. M. The phylogenetic position of the Cretaceous birds of Argentina: *Enantiornithes* and *Patagopteryx deferrariisi*. *Courier Forschungsinst. Senckenberg* **181**, 55–63 (1995).
- Chiappe, L. M. The first 85 million years of avian evolution. *Nature* **378**, 349–355 (1995).
- Sanz, J. *et al.* An Early Cretaceous bird from Spain and its implications for the evolution of avian flight. *Nature* **382**, 442–445 (1996).
- Ostrom, J. H. *Archaeopteryx* and the origin of birds. *Biol. J. Linn. Soc.* **8**, 91–182 (1976).
- Gauthier, J. A. Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* **8**, 1–46 (1986).
- Wilhelm de Moesbach, P. E., Rusca, W. M., Völet, A. & Suárez, E. *Nuevo diccionario Mapuche-Español*, 1–416 (Siringa Ediciones, Buenos Aires, 1993).
- Cruz, C. E., Condat, P., Kozłowski, E. & Manceda, R. Análisis estratigráfico secuencial del Grupo Neuquén (Cretácico superior) en el valle del Río Grande, Provincia de Mendoza. *I Congr. Arg. Hidrocarb.* **2**, 689–714 (1989).
- Novas, F. E. Alvarezsauridae, Cretaceous basal birds from Patagonia and Mongolia. *Mem. Queens. Mus.* **39**, 675–702 (1996).
- Novas, F. E. Anatomy of *Patagonykus puertai* (Theropoda, Avialae, Alvarezsauridae), from the Late Cretaceous of Patagonia. *J. Vert. Paleont.* **17**, 137–166 (1997).
- Novas, F. E., Cladera, G. & Puerta, P. New theropods from the Late Cretaceous of Patagonia. *J. Vert. Paleont.* **16**, 56A (1996).
- Wellnhofer, D. Das fünfte Skelettexemplar von *Archaeopteryx*. *Palaeontographica (A)* **147**, 169–216 (1974).
- Wellnhofer, P. Das siebte Exemplar von *Archaeopteryx* aus den Solnhofener Schichten. *Archaeopteryx* **11**, 1–48 (1993).
- Psihoyos, L. *Hunting Dinosaurs* (Random House, New York, 1994).
- Jenkins, F. The evolution of the avian shoulder joint. *Am. J. Sci.* **293A**, 253–267 (1993).
- Ostrom, J. H. Osteology of *Deinonychus antirrhopus*, an unusual theropod dinosaur from the Lower Cretaceous of Montana. *Peabody Mus. Nat. Hist. Bull.* **30**, 1–165 (1969).
- Martin, L. in *Origin of the Higher Groups of Tetrapods* (eds Schultze, H.-D. & Trueb, L.) 485–540 (Comstock Publishing Associates, Ithaca, 1991).
- Ostrom, J. H. The pectoral girdle and forelimb function of *Deinonychus* (Reptilia: Saurischia): a correction. *Postilla* **165**, 1–11 (1974).
- Walker, C. A. New subclass of birds from the Cretaceous of South America. *Nature* **292**, 51–53 (1981).
- Ostrom, J. H. On a new specimen of the Lower Cretaceous theropod dinosaur *Deinonychus antirrhopus*. *Breviora* **439**, 1–21 (1976).
- Chiappe, L. M. Late Cretaceous birds of southern South America: anatomy and systematics of *Enantiornithes* and *Patagopteryx deferrariisi*. *Münchner Geowiss. Ab. (A)* **30**, 203–244 (Verlag Dr Friedrich Pfeil, 1996).
- Martin, L. D. & Tate, J. The skeleton of *Baptornis advenus* (Aves: Hesperornithiformes). *Smith. Contrib. Paleobiol.* **27**, 35–66 (1976).
- Marsh, O. C. *Odontornithes: a monograph on the extinct toothed birds of North America*. *United States Geological Exploration of the 40th Parallel*, 1–201 (Government Printing Office, Washington, 1880).
- Ostrom, J. H. *Archaeopteryx* and the origin of flight. *Q. Rev. Biol.* **49**, 27–47 (1974).
- Gauthier, J. A. & Padian, K. in *The Beginnings of Birds* (eds Hecht, M., Ostrom, J., Viohl, G. & Wellnhofer, P.) 185–197 (Freunde des Jura-Museums Eichstätt, Eichstätt, 1995).
- Cagle, G. Balda, R. & Willis, W. The physics of leaping animals and the evolution of preflight. *Am. Nat.* **121**, 455–476 (1983).
- Ostrom, J. H. The cursorial origin of avian flight. *Mem. Calif. Acad. Sci.* **8**, 73–81 (1986).

Acknowledgements. We thank L. Chiappe, P. Currie, F. A. Jenkins and L. Witmer for review of the manuscript. The specimen was discovered by A. Scanferla, and prepared by P. Puerta and S. Reuil. We thank C. Schaff for access to specimens housed at MCZ, Massachusetts. Figures were drawn by J. González. Assistance in Neuquén was kindly offered by R. Coria (MCZ, Plaza Huincul). This research was supported by the National Geographic Society and The Dinosaur Society.

Correspondence and request for materials should be addressed to F.E.N. (e-mail: fernovas@muanbe.gov.ar).

Trade-off-invariant rules for evolutionarily stable life histories

Eric L. Charnov

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

Optimization models have been widely and successfully used in evolutionary ecology to predict the attributes of organisms¹⁻⁸. Most such models maximize darwinian fitness in the face of trade-offs and constraints; the numerical results usually depend on the exact form of the trade-offs or constraints. But not always⁹: for example, earlier work⁹ predicted that the optimal range in offspring size ought to show a -1 scaling with small litter size, independent of most details of the underlying offspring-survival/offspring-size trade-off relation. Here I report that in non-growing (stationary), age-structured populations, three major life-history attributes (age at first breeding, size of an offspring in large litters, and reproductive effort) are likely to evolve to equilibrium values that satisfy a universal numerical rule; the underlying trade-off will have a slope of -1 at the optimum, independent of most other aspects of the trade-off. Each of these three attributes can be viewed as an allocation problem between just two alternatives; the trade-off is then between having more of one alternative and less of the other. The slope of the trade-off is simply the slope of the curve of allowed combinations of the two alternatives. The theory predicts that natural selection will push to an equilibrium where the slope is always -1 . The economic structure is the same as that which underlies evolution of the sex ratio where the two alternatives are sons and daughters^{2,10}.

Consider an economic problem in the allocation between two alternatives X_1 and X_2 in which allowed values must fall on (or within) a constraint or trade-off curve, as shown in Fig. 1. Further suppose that the utility function (fitness) maximized is the simple product $X_1 \cdot X_2$, which is equivalent to maximizing $\ln X_1 + \ln X_2$. Figure 1 is the standard graphical solution from economics, which plots curves of equal utility, or in our case fitness ($\ln X_1 + \ln X_2$ is constant). The optimum is where the highest equal-fitness curve just touches the trade-off curve. If $\ln(\text{fitness})$ equals $\ln X_1 + \ln X_2$, the optimum will always be where the trade-off curve has a slope of -1 , for any smooth trade-off curve. This is, of course, where the percentage increase in X_1 is just matched by the percentage decrease in X_2 (refs 2 and 10).

With fairly elementary economics, some evolutionary allocation problems can indeed be studied as two-dimensional (X_1, X_2) and yield darwinian fitness to be a simple product. The classic example is evolution of sex allocation² (sex ratio, sperm versus eggs for hermaphrodites, time as a male (female) for a sex changer), where autosomal inheritance makes fitness a simple product^{2,10}: gain-via-male times gain-via-female. It is this product structure for fitness combined with the two-dimensional allocation that makes sex allocation the most successful ESS (optimization) theory in evolutionary ecology²⁻⁶.

Are any other life-history problems naturally two-dimensional with fitness a product? Box 1 displays a surprisingly simple answer. R_0 , the 'net reproductive rate' for age-structured life histories, can always be written as a product of three demographic averages, aggregated variables, which neatly summarize the life history:

$$R_0 = S(\alpha) \cdot \bar{b} \cdot E(\alpha) \quad (1)$$

where S is the probability of surviving to the age of first breeding (α), $\bar{b}$ is the average rate of offspring production over the adult lifespan, and $E(\alpha)$ is the average length of the adult lifespan (that is,

beginning at age α). R_0 is a measure of darwinian fitness appropriate within stationary (non-growing) populations⁶. $R_0 \approx 1$, owing to density dependence for typical individuals, so the trade-offs are viewed as existing for individuals who deviate from the typical⁶. Thus, it is R_0 for a mutant which is the fitness measure of interest.

Because equation (1) holds for any age-structured life history, evolutionary rules that follow from it should be quite general; the trick is to make predictions in terms of the three aggregate variables: S , E and $\bar{b}$. I consider here three classic allocation problems in the evolution of life histories.

First, consider 'reproductive effort'. $S \cdot \bar{b}$ is the rate of production of female babies (alive at age α); E is the mother's reproductive lifespan. Denote $S \cdot \bar{b}$ as B . Beginning with Williams¹¹, it has been widely postulated that B should generally trade-off against E (refs 7, 8); effort devoted to offspring production (B) should decrease the mother's own reproductive lifespan. Although there are life-history models¹² that do not make this 'reproductive effort' trade-off assumption, it is a widely held idea^{7,8}. Usually the trade-off assumption is implemented on an age-by-age basis (following ref. 11), but nothing precludes working in terms of the aggregate variables B and E . Indeed, many existing models^{7,8} can be alternatively viewed in terms of these averages. Assume it is generally true that natural selection sets an optimal balance between B and E . As fitness (equation (1)) is the product of B and E , Fig. 1 tells us that the optimum will be the place on the $\ln B, \ln E$ trade-off surface where the slope is -1 : $\partial \ln B / \partial \ln E = -1$, for all age-structured life histories where R_0 can be used for fitness.

Second, suppose adult lifespan, $E(\alpha)$, is unrelated to how reproductive resources are divided among offspring; then $R_0 \propto S \cdot \bar{b}$, a product, and we have a classic offspring quality (S) versus number ($\bar{b}$) tradeoff (ch. 7 in ref. 7; refs 9, 13). At the optimum allocation $\partial \ln \bar{b} / \partial \ln S = -1$ on the $\ln S, \ln \bar{b}$ trade-off surface¹³. This -1 rule also follows for many other size-versus-number allocations¹³, where fitness is naturally expressed as gain per unit multiplied by the number of units. This -1 rule assumes maternal control over the allocation process. The rule will not hold if optimal investment per offspring is under offspring control; deviations from -1 may well be a useful signature of offspring control¹⁴.

Third, write equation (1) as $R_0 = S(\alpha) \cdot V(\alpha)$, where $V(\alpha)$ is $\bar{b} \cdot E(\alpha)$, the average production of offspring over the adult lifespan⁶. This is the classic trade-off with respect to age at maturity (α ; ref. 7, ch. 6); S goes down with α , while V goes up. As fitness is a product

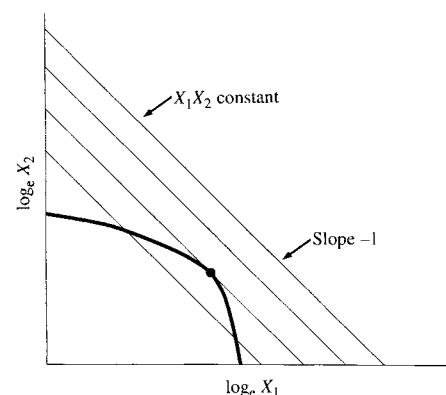


Figure 1 Economic optimization in two dimensions. Allowed combinations of two alternatives (X_1, X_2) are on (or within) the trade-off curve, here plotted for $\ln X_2$ versus $\ln X_1$, as a thick curve. If fitness is the simple product $X_1 \cdot X_2$, the optimum (or ESS) is found by plotting isofitness ($\ln X_1 + \ln X_2 = C$, a constant) curves and finding where the highest one just touches the trade-off curve. For all possible smooth trade-off curves, this yields a slope of -1 at the ESS because -1 is the slope of any isofitness curve: $\partial \ln X_2 / \partial \ln X_1 = -1$ always, for a function $\ln X_2 = C - \ln X_1$.

($S \cdot V$), the optimum is (again) where $\partial \ln S / \partial \ln V = -1$ on the $\ln S$, $\ln V$ trade-off surface. This result is derived, but not really understood, in ref. 6.

It seems that several classic problems in life-history evolution yield a universal trade-off slope of -1 at the optimum because aggregated fitness (Box 1) is naturally expressed as a product of two allocation alternatives (E versus B , $\bar{b}$ versus S , S versus V). Life histories are often treated as complex objects, with numerous possible age-dependent trade-offs^{7,8,15}. The approach here reduces them to just a few aggregate variables (S , $\bar{b}$, E)⁶. Although we sacrifice all information about age-dependent allocation decisions^{7,8}, we gain in finding general rules (hypotheses) about the shape of aggregate (average) trade-off surfaces at the equilibrium. I do not know of any data on trade-offs precise enough to test these differential-invariant predictions, particularly for the size/number or reproductive effort problems⁷. Theory for the optimal age at first reproduction using R_0 as a fitness measure often implicitly invokes the minus-one rule as an intermediate step in the prediction of attributes such as optimal adult body mass^{6,7,12}. Thus the minus-one rule is tested, at least qualitatively, whenever

Box 1 Fitness is a product

The 'net reproductive rate' R_0 is defined as $R_0 = \int_0^\infty l(x) \cdot b(x) dx$, and calculates the average number of daughters produced over a female's lifespan. $l(x)$ is the probability of being alive at age x ; $b(x)$ is the daughters produced at age x who are alive at independence from mother; $b > 0$ only if $x > \alpha$, the age at first birth, measured from independence. Now, write $b(x) = b(y)$ for $y = x - \alpha$ and denote $l(x)$ for $x > \alpha$ as $l(x) = S(\alpha) \cdot e^{-\phi(x-\alpha)} = S(\alpha) \cdot e^{-\phi(y)}$. (Notice that $\phi(y)$ is zero at $y = 0$ and is increasing with y ; $\partial(-\log l(x))/\partial y = \partial \phi(y)/\partial y$, so $\partial \phi/\partial y$ is the instantaneous mortality rate at age y .) $S(\alpha)$ is the chance of living from independence to α . R_0 can be written for this general life history as

$$R_0 = S(\alpha) \cdot \int_0^\infty b(y) \cdot e^{-\phi(y)} dy \quad (2)$$

Recall from the stable age distribution theory that the proportion of the breeding lifespan spent between ages y and $y + dy$ (the probability density function for the adult ages) is given by

$$\frac{e^{-\phi(y)}}{\int_0^\infty e^{-\phi(y)} dy} dy$$

Now, multiply equation (2) by (1), written as

$$\frac{\int_0^\infty e^{-\phi(y)} dy}{\int_0^\infty e^{-\phi(y)} dy}$$

to yield

$$R_0 = S(\alpha) \left[\int_0^\infty b(y) \cdot \frac{e^{-\phi(y)}}{\int_0^\infty e^{-\phi(y)} dy} dy \right] \left(\int_0^\infty e^{-\phi(y)} dy \right) \quad (3)$$

$S(\alpha)$ is the chance of living to reproduce at age α ; while the term in square brackets is simply $\bar{b}$, the average rate of production of offspring over the reproductive adult life, and the term in curved brackets is simply $E(\alpha)$, the expectation of further life at age α , the average length of the adult lifespan. So equation (3) is really

$$R_0 = S(\alpha) \cdot \bar{b} \cdot E(\alpha) \quad (4)$$

Equation (4) applies to any age structured life history; R_0 is the simple product of three aggregated terms, each an average. For equation (4) to be used as a fitness measure, the population must not be growing. This makes $R_0 = 1$ for typical individuals owing to density dependence⁶. But mutant individuals may have their own $R_0 \neq 1$, and it is fitness and trade-offs for mutants which are discussed here. Thus we use R_0 , equation (4), as a fitness measure, with the condition that it must equal unity at the optimum, when mutants are the same as typical¹.

these predictions work out. See, for example, the successful prediction of the heights and slopes of the between-species lifespan allometries for various mammal groups (ref. 6, p. 96; ref. 16). Of course, there are many qualitative (and a few quantitative) tests of product maximization for evolution of sex allocation^{2,6}.

The procedure described here is to reduce darwinian fitness to a function of a few aggregate variables, hoping to find a general form for fitness (here a product) which then yields general rules for the equilibrium. The trick is worth trying for other problems in phenotypic evolution. Economists often use this procedure and, indeed, production or utility function in the form of products are common; $X_1 \cdot X_2$, or more generally $X_1^A \cdot X_2^D$ (see any advanced text on price theory; A and D are > 0 and scale the relative productive value of inputs to X_1 and X_2 , respectively). Then products like that of equation (1) may often characterize fitness in non-growing populations with reproductive structure even more complex than simply age. □

Received 30 January; accepted 3 April 1997.

1. Parker, G. A. & Maynard Smith, J. Optimality theory in evolutionary biology. *Nature* **348**, 27–33 (1990).
2. Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, NJ, 1982).
3. Seger, J. & Stubblefield, J. W. in *Adaptation* 93–123 (Academic, New York, 1996).
4. Bulmer, M. G. *Theoretical Evolutionary Ecology* (Sinauer, Sunderland, MA, 1994).
5. Godfray, H. C. J. *Parasitoids* (Princeton Univ. Press, NJ, 1994).
6. Charnov, E. L. *Life History Invariants* (Oxford Univ. Press, 1993).
7. Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, 1992).
8. Charlesworth, B. *Evolution in Age-Structured Populations*. (Cambridge Univ. Press, 1994).
9. Charnov, E. L. & Downhower, J. F. A trade-off-invariant life-history rule for optimal offspring size. *Nature* **376**, 418–419 (1995).
10. MacArthur, R. H. in *Theoretical and Mathematical Biology* (eds Waterman, T. H. & Morowitz, H.) 388–397 (Blaisdell, New York, 1965).
11. Williams, G. C. Natural selection, the cost of reproduction, and a refinement of Lack's principle. *Am. Nat.* **100**, 687–690 (1966).
12. Charnov, E. L. Evolution of life history variation among female mammals. *Proc. Natl Acad. Sci. USA* **88**, 1134–1137 (1991).
13. Lloyd, D. G. Selection of offspring size at independence and other size-versus-number strategies. *Am. Nat.* **129**, 800–817 (1987).
14. Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford Univ. Press, 1997).
15. Stearns, S. C. Tradeoffs in life history evolution. *Funct. Ecol.* **3**, 259–268 (1989).
16. Charnov, E. L. & Berrigan, D. Why do female primates have such long lifespans and so few babies? *Evol. Anthropol.* **1**, 191–194 (1993).

Acknowledgements. K. Hawkes, G. Parker, G. Getty and S. Proulx read and improved the arguments.

Correspondence and requests for materials should be addressed to the author.

Molecular basis of symbiosis between *Rhizobium* and legumes

Christoph Freiberg, Rémy Fellay*, Amos Bairoch†, William J. Broughton*, André Rosenthal & Xavier Perret*

Institut für Molekulare Biotechnologie, Abteilung Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany

*Laboratoire de Biologie Moléculaire des Plantes Supérieures, Université de Genève, 1 ch. de l'Impératrice, 1292 Chambésy/Genève, Switzerland

†Biochimie Médicale, Centre Médical Universitaire, Université de Genève, 1 rue Michel-Servet, 1211 Genève 4, Switzerland

Access to mineral nitrogen often limits plant growth, and so symbiotic relationships have evolved between plants and a variety of nitrogen-fixing organisms. These associations are responsible for reducing 120 million tonnes of atmospheric nitrogen to ammonia each year. In agriculture, independence from nitrogenous fertilizers expands crop production and minimizes pollution of water tables, lakes and rivers. Here we present the complete nucleotide sequence and gene complement of the plasmid from *Rhizobium* sp. NGR234 that endows the bacterium with the ability to associate symbiotically with leguminous plants. In conjunction

with transcriptional analyses, these data demonstrate the presence of new symbiotic loci and signalling mechanisms. The sequence and organization of genes involved in replication and conjugal transfer are similar to those of *Agrobacterium*, suggesting a recent lateral transfer of genetic information.

Many dissimilar organisms live in close association with one another. Nitrogen-fixing symbioses, in which auxotrophs exchange carbohydrates for organic nitrogen with diazotrophs, dominate numerous ecological niches. Members of the plant family Leguminosae, in association with the soil bacteria *Azorhizobium*, *Bradyrhizobium* and *Rhizobium*, are responsible for most of the nitrogen fixed biologically. Plant roots excrete a variety of substances, some of which (especially flavonoids) coordinate the expression of rhizobial nodulation (*nod*) genes¹. In turn, the Nod proteins direct synthesis of lipochito-oligosaccharidic Nod factors, which initiate nodule formation and allow rhizobia to enter the plant^{2,3}. Continued exchange of symbiotic signals is necessary for nitrogen fixation⁴.

As a general rule, symbiotic genes are plasmid borne in *Rhizobium* species and are located on the chromosome in *Azo*(*Brady*)*rhizobium* strains. To study the molecular control of broad host range in associations between legumes and *Rhizobium*, we use the fast-growing *Rhizobium* sp. NGR234 (ref. 5). This bacterium nodulates more than 110 genera of legumes, as well as the non-legume *Parasponia andersonii* (S. G. Pueppke and W.J.B., unpublished data), and possesses a large plasmid (pNGR234a) that carries most symbiotic determinants⁶. Using dye terminators and a thermostable sequenase⁷, we sequenced 20 cosmids from the canonical ordered library⁸, selected to cover pNGR234a.

The symbiotic replicon is 536,165 base pairs long (92% of the genome of *Mycoplasma genitalium*⁹). A total of 416 open reading frames (ORFs) were predicted to encode proteins (Fig. 1, Table 1), 139 of which show no similarity to any known protein. An additional 67 gene fragments were detected that seem to be remnants of functional genes. Many of these have clearly been disrupted by mobile elements. Collectively, potential genes and gene fragments make up 418 kilobases (78% of the replicon). This gene density is slightly lower than that of other bacterial genomes such as *Bacillus subtilis*, *Escherichia coli* and *Haemophilus influenzae*¹⁰ (85–92%; I. Moszer, personal communication), mostly because of the numerous insertion sequences and of their recombinatorial nature. No genes essential to transcription, translation or primary metabolism were found, which is consistent with the observation that NGR234 can be cured of its plasmid, giving strain ANU265 (ref. 11).

In total, 85 proteins belong to families that are represented more than once, even after discounting the many insertion-sequence encoded proteins or those involved in transposition, integration and recombination. There are some large families, including the five members of the short-chain dehydrogenase/reductases¹², one of which (y4vI) contains two homologous domains; four complete and one partial ABC-type transporter operons that each encode at least one ABC-type permease and ABC-type ATP-binding proteins; four cytochrome P450s; and three members of the peptidase family S9A.

Highest similarities were found with proteins of other rhizobia and agrobacteria. Only two eukaryotic proteins are highly homologous to pNGR234a counterparts. A glutamate dehydrogenase (y4uF) is significantly more closely related to the mitochondrial form than to that of prokaryotes or to archaea. As mitochondrial proteins are thought to originate from ancestral bacteria, y4uF could thus represent a new subclass of prokaryotic glutamate dehydrogenases. A small protein (y4sK), which belongs to a family (Prosite:PDO00838) of uncharacterized proteins found in a variety of prokaryotes and eukaryotes, is more closely related to the known mammalian members of this family than to that of other prokaryotes.

Replication of pNGR234a probably begins at *oriV*, which is

located within the intergenic sequence between the *repC* (y4cI)- and *repB* (y4cJ)-like genes. This locus is highly similar (69–71%) to the origins of replication of *Agrobacterium* and *Rhizobium* plasmids (Fig. 2). In *Agrobacterium*, these intergenic sequences are the determinants of incompatibility. RepABC of pNGR234a are 40–60% identical to those of pTiB6S3 (a Ti-plasmid), pRiA4b (an Ri-plasmid) and pRL8JI (a cryptic *Rhizobium* plasmid). A 12-bp portion of the origin of transfer (*oriT*) is identical to that of pTiC58 of *A. tumefaciens*, and highly similar to those of RSF1010 (*E. coli*) and pTF1 (*Thiobacillus ferrooxidans*). This sequence corresponds to the *oriT* of plasmids containing the 'Q-type nick-region' (Fig. 3). Adjacent to this cluster are another 21 predicted genes that are homologous to the conjugal transfer genes of *Agrobacterium* Ti-plasmids. By marking the *nodD2* gene (y4xH), we were able to demonstrate high-frequency conjugal transfer of pNGR234a into ANU265. Conjugal transfer of Ti plasmids in *Agrobacterium* is controlled by a family of *N*-acyl-L-homoserine lactone auto-inducers^{13,14}. By using established methods¹⁴, we found similar molecules, which interact with the *Agrobacterium traR* gene product, in supernatants of NGR234 cultures.

Carbohydrates are not only constituents of the rhizobial cell wall: they are also morphogens. Short, *N*-acylated tri- to pentamers of *N*-acetyl-D-glucosamine (Nod factors) trigger nodulation responses in homologous legumes at very low concentrations¹⁵. Thus elements of the biosynthetic pathways leading to cell walls or to Nod factors are common. Most differences are found in the later stages of the pathways that yield specific cell-wall components or Nod factors. Yet even here, the distinction between structural and symbiotic carbohydrates is blurred. Specific extracellular polysaccharides (EPS) are required for nitrogen fixation in certain legume–*Rhizobium* associations¹⁶.

As befits a symbiotic replicon, only 10 ORFs with similarities to polysaccharide synthesis genes (*sensu stricto*) are plasmid borne (Table 1). Sequences homologous to *exo* genes are clearly located on the chromosome (X.P. and V. Viprey, unpublished data). Although loci with weak homologies to *nod*-box::*psiB* of *R. leguminosarum*, and *exoX* of *R. meliloti* exist on pNGR234a (y4iR, and y4xQ respectively), these are regulatory rather than structural.

Except for *nodE*, *nodG* and *nodPQ*, which are on the chromosome, most Nod-factor biosynthetic genes are plasmid borne, and are regulated by four transcriptional regulators of the *lysR* family: *nodD1* (y4aL), *syrM1* (y4pN), *nodD2* (y4xH) and *syrM2* (y4zF). Most *nod* genes share a conserved promoter sequence called the *nod* box. NodC (an *N*-acetylglucosaminyltransferase), which is the first committed enzyme in the Nod-factor biosynthetic pathway, is part of the *hsnIII* locus: *nodABC*/*hslO* and *noeE* are responsible for synthesis of the core Nod factor as well as the adjunction of 3-(or 4)-*O*-carbomoyl-, 2-*O*-methyl-, and 4-*O*-sulphate¹⁷ groups, respectively (Fig. 4). The *hsnI* locus encodes enzymes involved in fucosylation of NodNGR factor¹⁸. NodS and NodU (*hsnII*) *N*-methylate and 6-*O*-carbomoylate NodNGR factors respectively¹⁹, whereas *noII* is probably an acetyltransferase.

We found 12 additional *nod* box-like sequences. Among these, two are upstream of the transcriptional regulators, *syrM2* and y4xI, and three (y4hM, *fixF* and y4iR) control expression of genes involved in polysaccharide metabolism. Transcription analysis suggests that at least 11 of the 17 putative *nod* boxes are symbiotically active (Fig. 1). Many *nod* box-dependent genes are under the direct control of NodD1, but y4wM is modulated by NodD2 (data not shown). The presence of both a *nod* box and a NifA- σ^{54} -type promoter upstream of y4vC suggests a new system of transcriptional regulation involving both flavonoid induction and expression in nodules (Fig. 5).

In contrast to the mosaic structure of the *nod* loci, pNGR234a contains a single large cluster of 43 *nif* and *fix* genes, including *nifA* (y4uN), which encodes a σ^{54} -dependent regulator. NifA contains two highly conserved regions, a carboxy-terminal part involved in

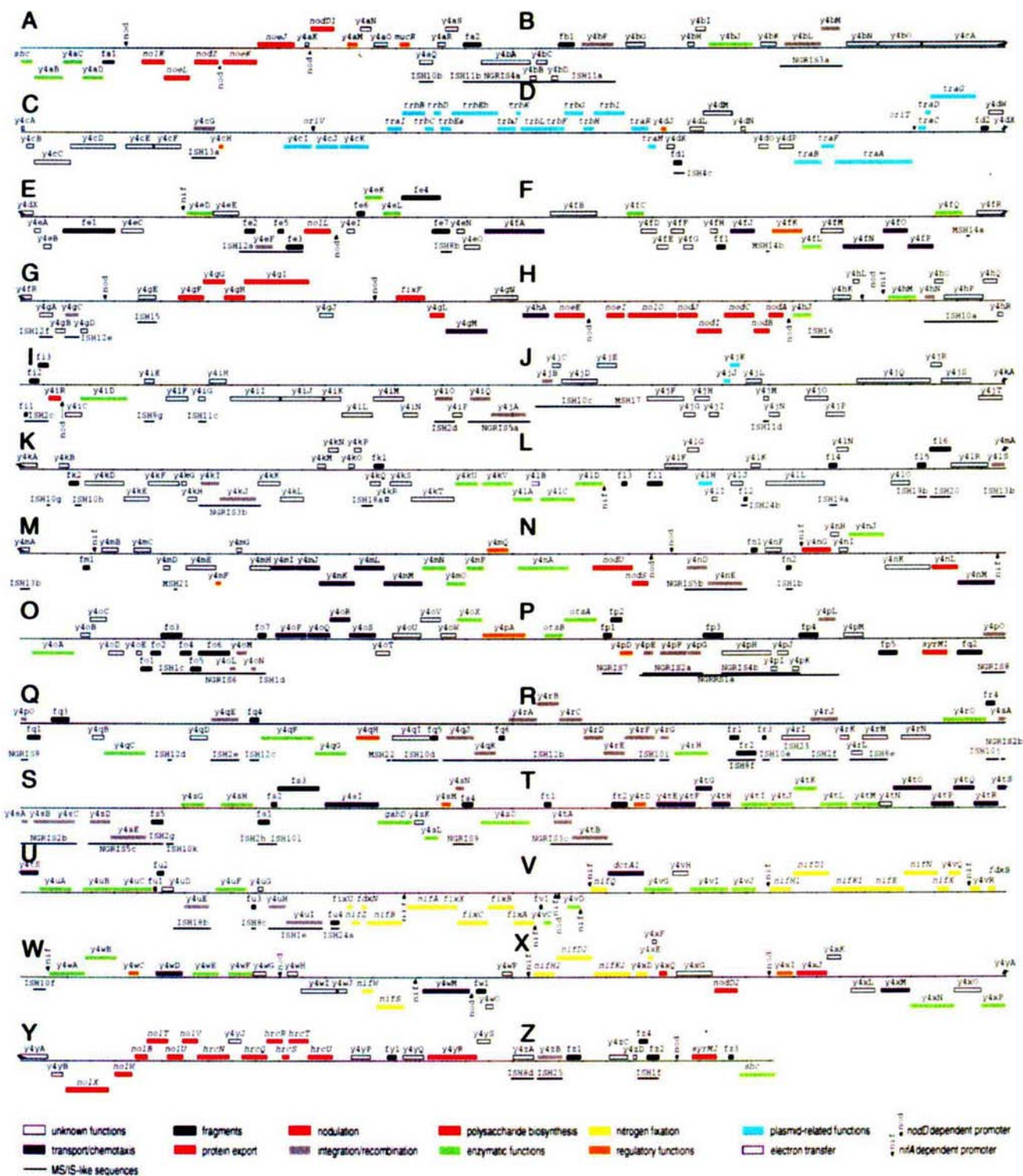


Figure 1 Genetic map of the symbiotic replicon pNGR234a. Each line represents 42 kb. ORF names (for example, y4aB) or the names of genes (for example, *shc*) correspond to those in Table 1. The third letter in each ORF name corresponds to one of 26 segments (A to Z), each of which is 21 kb in length except segment Z, which comprises 11,125 bp. The fourth letter of the ORF name indicates the position of the ORF within each segment (for example, y4gI is located on segment G between y4gH and y4gJ). Genes positioned on top of each line are transcribed from left to right, whereas those placed below the lines are encoded by the complementary strand. Colour-coded boxes represent putative coding regions

(and gene fragments) grouped according to their presumed functions, except for those involved in nodulation, polysaccharide synthesis, nitrogen fixation, and plasmid-related functions, which are grouped according to their phenotypic class (see Table 1). Positions of the origins of replication and transfer are marked *oriV* and *oriT*, respectively. Regions homologous to consensus sequences of *nod*-box- and *NifA*- σ^{54} dependent promoters are marked with a black arrowhead followed by *nod* and *nif*, respectively. Positions of insertion- and mosaic-sequences are annotated with lines (for example, ISH10b, NGRIS4a or MSH22).

DNA binding, and a central domain that interacts with an alternative sigma factor (σ^{54}) of the RNA polymerase. Mutation of *rpoN* (which encodes σ^{54}) causes a Fix⁻ phenotype on NGR234 hosts²⁰. Genes involved in the synthesis of the MoFe-nitrogenase complex are also present, including two identical copies of the *nifKDH* genes (*nifHDK1* and *nifHDK2*)²¹. Other *nif* and *fix* genes are involved in elaboration of the electron-transport complex (*fixA*, *fixB*), of various cofactors required for nitrogen fixation (*fixC*, *nifB*, *nifE*, *nifN*) and in the syntheses of ferredoxins (*fdxB*, *fdxN* and *fixX*). Although not directly involved in the fixation process, mutation of

the plasmid-borne copy of *dctA* (*dctA1*, *y4vF*) also impairs nitrogen fixation²².

In addition, 17 new genes expressed in nodules are part of this *nif* and *fix* cluster (*y4vC* to *y4vJ*, with the exception of *dctA1*, *y4wA* to *y4wG*, *y4wI*, *y4wJ* and *y4xQ*; see Fig. 5). The predicted functions of six of these (*y4vD*, *nifQ*, *y4vG*, *y4vI*, *y4vJ* and *y4wF*) suggest a role in electron transport and oxidation-reduction, but another five are not homologous to any database entry. It thus seems likely that most of the proteins necessary for bacteroid development and synthesis of the nitrogen-fixing complex are coded by pNGR234a, although



Figure 2 Multiple alignments of the nucleotide sequence of the replication origins of: the Ri plasmid of *Agrobacterium rhizogenes* (pRI4b); the symbiotic replicon of NGR234 (pNGR234a); the Ti plasmid of *A. tumefaciens* BS63 (pTiBS63); and pRLJ of *R. leguminosarum* bv. *leguminosarum* (pRLJ). Gaps introduced to give the best sequence alignments are marked with hyphens.

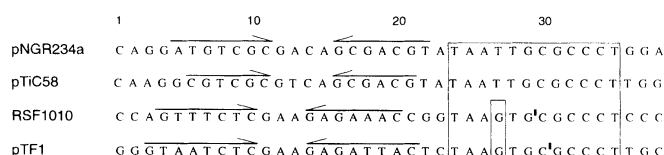


Figure 3 Multiple DNA sequence alignments of loci containing the origin of transfer of: the symbiotic plasmid of NGR234 (pNGR234a); the Ti plasmid of *A. tumefaciens* C58 (pTiC58); a mobilizable plasmid of *E. coli* (RSF1010) and a mobilizable plasmid of *Thiobacillus ferrooxidans* (pTFI). Major conserved nucleotide residues are boxed. Known 'nick' sites corresponding to the nucleotide positions where the specific plasmid strand is cleaved are marked with black boxes. Sequence features in the trailing portion of the *oriT* sites include inverted repeats, which are marked with horizontal arrows.

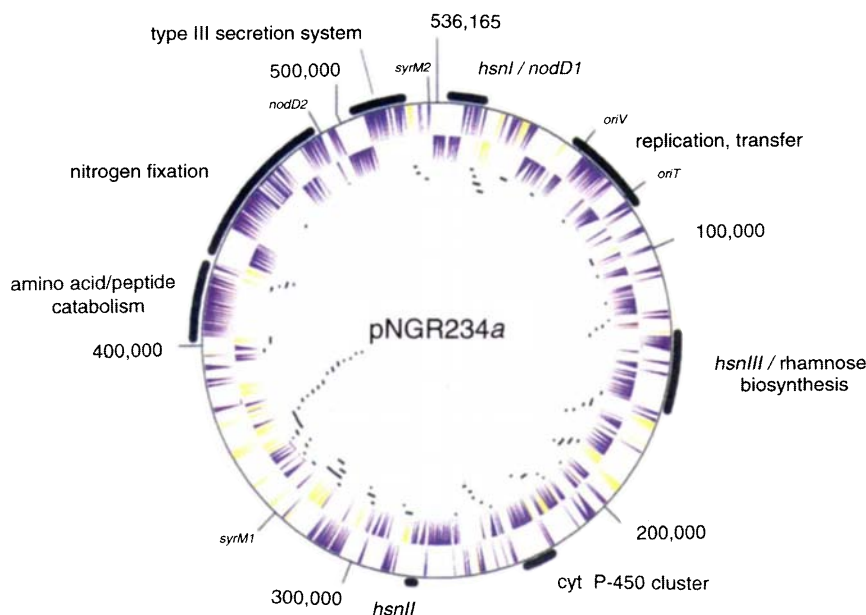


Figure 4 Circular representation of pNGR234a. Outer and inner concentric circles indicate coding regions identified on the plus and minus strands, respectively. Within these predicted genes, those coloured in yellow correspond to ORFs that belong to insertion-like elements. Thin concentric black lines represent mosaic

sequences as well as complete or partial insertion sequence-like repeats. Major gene clusters are highlighted as thick black concentric lines (for example, *hsn*, host specificity of nodulation loci).

Table1 Functional and phenotypic classification of the predicted proteins encoded by pNGR234a

general statistics		proteins sorted according to phenotypic classes		other proteins	
		functional class	number of proteins	functional class	number of proteins
total number of predicted proteins	416	nodulation		various enzymatic functions	
mean length of the putative proteins (smallest: 43 aa, largest: 1197 aa)	308 aa	enzymes		59	
putative subcellular locations of the predicted proteins:		transporters		transport	
- cytoplasmic	329	transcriptional regulators		29	
- transmembrane (inner membrane)	62	export proteins		protein export/secretion	
- periplasmic	20	unknown functions		8	
- outer membrane	2	poly-/oligosaccharide synthesis		chemotaxis	
- lipoprotein attached to a membrane	3	enzymes		2	
number of proteins grouped into functional classes		transcriptional regulators		electron transfer	
	234	nitrogen fixation		transcriptional regulation	
number of unknown proteins:		enzymes		14	
- with database homolog	43	electron transfer proteins		integration/recombination	
- without any database homolog	139	transcriptional regulators		44	
		unknown functions			
		plasmid-related functions/cell development			
		enzymes			
		export proteins			
		transcriptional regulators			
		plasmid control/cell development proteins			
		12			

ORF name	gene name	description of the predicted protein							
nodulation									
Nod-factor biosynthesis/modification/transport									
y4aF	<i>noiK</i>	pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase		y4uP	<i>fixC</i>	nitrogenase cofactor biosynthesis		y4eD	pu. phosphodiesterase
y4aG	<i>noeL</i>	GDP-mannose 4,6-dehydratase		y4vA	<i>fixB</i>	electron transfer fixAB complex		y4eK	pr. short chain oxidoreductase
y4aH	<i>nodZ</i>	fucosyltransferase		y4vB	<i>fixA</i>	electron transfer fixAB complex		y4eL	pr. short chain oxidoreductase
y4aI	<i>noeK</i>	phosphomannomutase		y4vE	<i>nifQ</i>	pu. nitrogenase Mo cofactor processing		y4fC*	pu. monooxygenase
y4aJ	<i>noeJ</i>	mannose-1-phosphate guanylyltransferase		y4vK	<i>nifH1</i>	nitrogenase Fe protein; id. tq y4xA		y4fL	member of inositol monophosphatase family
y4eH	<i>noiL</i>	sim. to <i>R. loti</i> acetyltransferase NoIL		y4vL	<i>nifD1</i>	nitrogenase MoFe protein alpha subunit; id. to y4xB		y4fQ	'ROK' family member
y4hB	<i>noeE</i>	transfer of activated sulfate to fucose		y4vM	<i>nifK1</i>	nitrogenase MoFe protein beta subunit; id. to y4xC		y4gN	sim. to <i>V. anguillarum</i> VirA
y4hC	<i>noeI</i>	not yet characterized step in Nod-factor synthesis		y4vN	<i>nifE</i>	nitrogenase MoFe cofactor synthesis		y4hJ*	pu. oxidoreductase; sim. to N-terminus of anaerobic coproporphyrinogenIII oxidases
y4hD	<i>noiO</i>	O-carbamoylation of Nod-factors		y4vO	<i>nifN</i>	nitrogenase MoFe cofactor synthesis		y4hM	pu. oxidoreductase; Low similarity to <i>Z. mobilis</i> glucose-fructose oxidoreductase
y4hE	<i>nodJ</i>	pr. ABC transporter permease		y4vP	<i>nifX</i>	unknown function		y4iD	pr. monooxygenase
y4hF	<i>nodI</i>	pr. ABC transporter ATP-binding protein		y4vQ		sim. to <i>nifX-nifW</i> intergenic ORF in nitrogen-fixing bacteria		y4kU	pr. geranyltransferase
y4hG	<i>nodC</i>	N-acetylglucosaminyltransferase		y4vR		sim. to N-terminus of NifH		y4kV	pr. cytochrome P450
y4hH	<i>nodB</i>	chito oligosaccharide deacetylase		y4vS	<i>fdxB</i>	4Fe-4S ferredoxin		y4kX	pr. short chain oxidoreductase
y4hI	<i>nodA</i>	N-acetyltransferase		y4vT	<i>nifW</i>	unknown function		y4kY	pr. cytochrome P450
y4hJ	<i>nodU</i>	6-O-carbamoylation of Nod-factors		y4vU	<i>nifS</i>	class-5 aminotransferase		y4kZ	pr. cytochrome P450
y4hK	<i>nodS</i>	methyltransferase involved in Nod-factor biosynthesis		y4vV	<i>nifH2</i>	nitrogenase Fe protein; id. to y4vK		y4kA	pr. TPP enzyme (transketolase family; C-terminus)
y4hL				y4vW	<i>nifD2</i>	nitrogenase MoFe protein alpha subunit; id. to y4vL		y4kM*	pr. TPP enzyme (transketolase family; N-terminus)
y4hM				y4vX	<i>nifK2</i>	nitrogenase MoFe protein beta subunit; id. to y4vM		y4kN*	pr. short chain oxidoreductase
y4hN				y4vY		sim. to <i>nifX-nifW</i> intergenic ORF in nitrogen-fixing bacteria		y4kP	pr. peptidase (S9A family)
y4hO				y4vZ		sim. to <i>nifX-nifW</i> intergenic ORF in nitrogen-fixing bacteria		y4kR	pr. GMC-type oxidoreductase
transcriptional regulation									
y4aL	<i>nodD1</i>	LysR family H-T-H regulator		y4xP				y4oX	pr. NAD-dependent oxidoreductase
y4pN	<i>syfM1</i>	LysR family H-T-H regulator		y4xQ				y4pB	pr. trehalose-phosphatase
y4xH	<i>nodD2</i>	LysR family H-T-H regulator		y4xR				y4pC	pr. α,α -trehalose-phosphate synthase
y4zF	<i>syfM2</i>	LysR family H-T-H regulator		y4xE				y4qC	sim. to <i>E. coli</i> pMCC07 microcin biosynthesis protein MccB
protein secretion/unknown functions									
y4yC	<i>noiX</i>	unknown function		y4yD				y4qF	pr. peptidase (S9A family)
y4yD	<i>noiW</i>	secretion system protein (PulD family)		y4yE				y4gH	class-3 aminotransferase
y4yE	<i>noiB</i>	unknown function		y4yF				y4rN	pu. ligase; sim. to biotin carboxylase
y4yF	<i>noiT</i>	unknown function		y4yG				y4rO	sim. to C-terminus of histidinol-phosphate aminotransferase
y4yG	<i>noiU</i>	unknown function		y4yH				y4sG	pu. ligase; sim. to D-alg, D-alg ligases
y4yH	<i>noiV</i>	unknown function		poly-/oligosaccharide synthesis					
y4gF		pr. dTDP-glucose 4,6-dehydratase		y4gG		pr. dTDP-4-dehydrothamnose reductase		y4sH	pu. cell wall compound biosynthesis protein; sim. to <i>B. anthracis</i> CapA
y4gH		pr. glucose-1-phosphate thymidyltransferase		y4gI		pu. enzyme involved in polysaccharide biosynthesis; sim. to <i>Myx. xanthus</i> O-antigen RfbC		y4sJ	pr. succinate-semialdehyde dehydrogenase
y4gJ		pu. enzyme involved in polysaccharide biosynthesis; sim. to <i>Myx. xanthus</i> O-antigen RfbC		y4gK	<i>fixF</i>	pu. enzyme involved in polysaccharide biosynthesis; Low similarity to <i>E. coli</i> KpsS		y4sL	pu. oxidoreductase; sim. to C-terminus of <i>E. coli</i> D-amino acid dehydrogenase small subunit
y4gL		pr. dTDP-4-dehydrothamnose 3,5-epimerase		y4gM		possibly involved in polysaccharide biosynthesis; sim. to <i>R. leguminosarum</i> PsiB		y4sO	pr. peptidase (S9A family)
y4iR		possibly involved in polysaccharide biosynthesis; sim. to <i>R. leguminosarum</i> PsiB		y4gN		pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase		y4tI	pu. peptidase (M40 family)
y4nG		pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase		y4nL		pu. NAD-dependent nucleotide sugar epimerase/dehydrogenase		y4U	pu. threonine dehydratase
y4xQ		pu. exopolysaccharide production repressor; sim. to Rhizobium ExoX		y4yJ		pu. enzyme involved in polysaccharide biosynthesis; Low similarity to <i>E. coli</i> KpsS		y4tK	pu. cyclodeaminase
nitrogen fixation									
y4uJ	<i>fixU</i>	unknown function		y4uK	<i>nifZ</i>	unknown function		y4tL	pu. peptidase/hydrolase (M24 family)
y4uL	<i>fdxN</i>	4Fe-4S ferredoxin		y4uM	<i>nifB</i>	nitrogenase cofactor biosynthesis		y4tM	pu. cell wall comp. biosynthesis protein; sim. to <i>B. anthracis</i> CapA
y4uN	<i>nifA</i>	nitrogenase cofactor biosynthesis		y4uO	<i>nifA</i>	sigma54-dependent H-T-H regulator		y4uA	class-3 aminotransferase
y4uP	<i>fixX</i>	pr. 3Fe-3S ferredoxin		y4uQ				y4uB	pr. aldehyde dehydrogenase
plasmid-related functions/cell development									
y4cI		pu. replication protein; sim. to <i>A. rhizogenes</i> repC		y4cJ		pu. replication protein; sim. to <i>A. rhizogenes</i> repB		y4uF	pr. glutamate dehydrogenase
y4cK		pu. replication protein; sim. to <i>A. rhizogenes</i> repA		y4cL	<i>traI</i>	pr. autoinducer synthetase		y4vC	member of HesB/YadR/YthF family
y4cM	<i>traB</i>	pr. conjugal transfer protein (PulE family)		y4cN	<i>traC</i>	pr. conjugal transfer protein (export prot.)		y4vD	pu. redox enzyme; sim. to <i>C. boidinii</i> peroxisomal proteins A/B & <i>H. influenzae</i> H10572
y4cO	<i>traD</i>	pr. conjugal transfer protein (export prot.)		y4cP	<i>traE</i>	pr. conjugal transfer protein (export prot.)		y4vG	pr. cytochrome P450
y4cQ	<i>traEb</i>	pr. conjugal transfer protein (export prot.)		y4cR	<i>traF</i>	pr. conjugal transfer protein (export prot.)		y4vH	pr. short chain oxidoreductase
y4dA	<i>traJ</i>	pr. conjugal transfer protein (export prot.)		y4cS	<i>traA</i>	pr. conjugal transfer protein (export prot.)		y4vJ	pu. monooxygenase; sim. to LuxA/B (bacterial luciferase)
y4dB	<i>traK</i>	pr. conjugal transfer protein (export prot.)		y4dC	<i>traL</i>	pr. conjugal transfer protein (export prot.)		y4wA	pr. zinc peptidase (M16 family)
y4dD	<i>traF</i>	pr. conjugal transfer protein (export prot.)		y4dE	<i>traG</i>	pr. conjugal transfer protein (export prot.)		y4wB	pu. zinc peptidase (M16 family)
y4dF	<i>traG</i>	pr. conjugal transfer protein (export prot.)		y4dG	<i>traI</i>	pr. conjugal transfer protein (export prot.)		y4wE	class-2 aminotransferase
y4dH	<i>traR</i>	LuxR family H-T-H regulator		y4dI	<i>traM</i>	pu. repressor		y4wF	pu. monooxygenase; sim. to LuxA/B (bacterial luciferase)
y4dJ	<i>traP</i>	pu. repressor		y4dK	<i>traB</i>	pr. conjugal transfer protein			
y4dL	<i>traQ</i>	pr. conjugal transfer protein		y4dM	<i>traF</i>	pr. conjugal transfer protein			
y4dN	<i>traA</i>	pr. conjugal transfer protein		y4dO	<i>traC</i>	pr. conjugal transfer protein			
y4dP	<i>traD</i>	pr. conjugal transfer protein		y4dQ	<i>traD</i>	pr. conjugal transfer protein			
y4dR	<i>traG</i>	pr. conjugal transfer protein		y4dS	<i>traE</i>	pr. conjugal transfer protein			
y4jI		pu. plasmid stability protein; sim. to <i>Ps. syringae</i> StbC		y4dT	<i>traF</i>	pr. conjugal transfer protein			
y4jK		pu. plasmid stability protein; sim. to <i>Ps. syringae</i> StbB		y4dU	<i>traD</i>	pr. conjugal transfer protein			
y4iH		pu. cell filamentation protein; sim. to enterobacterial protein Fic		y4dV	<i>traG</i>	pr. conjugal transfer protein			
various enzymatic functions									
y4aA	<i>shc</i>	pr. squalene-hopene cyclase		y4aB		pr. phytoene synthase			
y4aC		sim. to phytoene synthase		y4aD		sim. to phytoene synthases			
y4aE		pu. peptidase (S2C family)		y4aF					

Table 1 continued

y4xN	Low similarity to <i>E. coli</i> pCOLV-K30 aerobactin synthase subunit lucC	y4iO*	pu. transposase (sim. to IS1111A/IS1328/IS1533 family)	y4iC	sim. to <i>M. tuberculosis</i> Mtcy373.06
y4xP	pu. cysteine synthase	y4iQ	pu. transposase-associated ATP-binding protein; id. to y4nD & y4sD (sim. to IS21/IS1162 family)	y4iL	pr. AAA-family ATPase
transport		y4jA	pu. transposase; id. to y4nE & y4sE (sim. to IS21/IS1162 family)	y4iR*	sim. to N-terminus of <i>Erw. herbicola</i> ORF6 in <i>crtE-crtX</i> region
probable sugar transport systems		y4jB	sim. to IS elements IS2/IS1312/IS866 proteins	y4kS	highly sim. to <i>Br. japonicum</i> hyp. protein
y4mI	pr. ABC transporter binding prot.	y4jI	pu. transposase-associated ATP-binding protein; id. to y4bM & y4tA (sim. to IS21/IS1162 family)	y4kT	highly sim. to <i>Br. japonicum</i> hyp. protein
y4mJ	pr. ABC transporter permease	y4kI	pu. transposase; id. to y4bL & y4tB (sim. to IS21/IS1162 family)	y4iL	Member of <i>E. coli</i> YegE/YhdA/YhjK/YjcC family
y4mK	pr. ABC transporter ATP-binding protein	y4iS	pu. integrase/recombinase 'resolvase-type'	y4iO	sim. to <i>Ps. syringae</i> avirulence protein
y4oP	pr. ABC transporter binding protein	y4nD	pu. transposase-associated ATP-binding protein; id. to y4iQ & y4sD (sim. to IS21/IS1162 family)	y4mB	sim. to <i>E. coli</i> yciD; <i>V. cholerae</i> OmpW & <i>Ps. oleovorans</i> AikL
y4oQ	pr. ABC transporter permease	y4nE	pu. transposase; id. to y4jA & y4sE (sim. to IS21/IS1162 family)	y4mE	sim. to <i>E. coli</i> HipA & <i>H. influenzae</i> HI0665
y4oR	pr. ABC transporter permease	y4pE	sim. to <i>R. fredii</i> RFRS9 and <i>Azo. xylinum</i> IS1268 ORF; id. to y4sA	y4nH	sim. to <i>E. coli</i> ethidium bromide resistance protein MvrC
y4oS	pr. ABC transporter ATP-binding protein	y4pF	pr. transposase; id. to y4sB (sim. to IS1111A/IS1328/IS1533 family)	y4rN	sim. to <i>M. tuberculosis</i> Mtcy50.24
probable aminoacid/peptide transport systems		y4pG	sim. to <i>A. xylinum</i> IS1268 ORFA; id. to y4sC	y4sK	pr. important cellular function (Yer057C/Yil051C/Yigf family)
y4tE	pr. ABC transporter binding protein	y4pL	pu. transposase-associated ATP-binding protein (sim. to IS21/IS1162 family)	y4wH	sim. to <i>A. tumefaciens</i> plasmid pTiA6 ORF in pinF2 region
y4tF	pr. ABC transporter permease	y4pO	pr. transposase (Mutator family)	y4yJ	sim. to <i>R. fredii</i> USDA257 ORF7
y4tG	pr. ABC transporter permease	y4qE	pr. transposase (sim. to IS1111A/IS1328/IS1533 family)	y4zC	sim. to <i>Ps. syringae</i> avirulence protein AvrPph3
y4tH	pr. ABC transporter ATP-binding protein	y4qI	pu. transposase (sim. to IS801)	unknown function	
y4tO	pr. ABC transporter binding protein	y4qK	pu. integrase/recombinase 'phage-type'	(no database homologue)	
y4tP	pr. ABC transporter permease	y4rA	pu. integrase/recombinase 'phage-type'	derived from IS element-like sequences	
y4tQ	pr. ABC transporter permease	y4rB	pu. integrase/recombinase 'phage-type'	y4bA	id. to y4pH
y4tR	pr. ABC transporter ATP-binding protein	y4rC	pu. integrase/recombinase 'phage-type'	y4bB	id. to y4pI
y4tS	pr. ABC transporter ATP-binding protein	y4rD	pu. integrase/recombinase 'phage-type'	y4bC	id. to y4pJ
other transport systems		y4rE	pu. integrase/recombinase 'phage-type'	y4bD	id. to y4pK
y4fJ	pu. porin; sim. to <i>R. leguminosarum</i> OMP111A	y4rF	pu. integrase/recombinase 'phage-type'	y4gA	
y4fN	pr. ABC transporter permease	y4rG	pu. integrase/recombinase 'phage-type'	y4gE*	
y4fO	pr. ABC transporter ATP-binding protein	y4rH	pu. integrase/recombinase 'phage-type'	y4iE*	
y4fP	pr. ABC transporter ATP-binding protein	y4rI	pu. integrase/recombinase 'phage-type'	y4iG*	
y4gM	pu. ionic transporter; sim. to <i>E. coli</i> ChaA	y4rJ	pu. integrase/recombinase 'phage-type'	y4iP*	
y4hA	pu. permease (<i>E. coli</i> YiaN/YgiK family)	y4sA	sim. to <i>R. fredii</i> RFRS9 and <i>Azo. xylinum</i> IS1268 ORF; id. to y4pE	y4jM*	
y4mL	pu. permease (SBR family 7)	y4sB	pr. transposase; id. to y4pF (sim. to IS1111A/IS1328/IS1533 family)	y4kQ*	
y4mM	pu. permease; sim. to <i>Az. caudatus</i> NoeC	y4sC	sim. to <i>A. xylinum</i> IS1268 ORFA; id. to y4pG	y4mA	
y4nM	C4-dicarboxylate transporter	y4sD	pu. transposase-associated ATP-binding protein; id. to y4iQ & y4nD (sim. to IS21/IS1162 family)	y4oL	
y4vF	<i>dctA1</i> permease-type protein; sim. to <i>R. meliloti</i>	y4sE	pu. transposase; id. to y4jA & y4nE (sim. to IS21/IS1162 family)	y4oM	
y4wD	MosC	y4sF	pu. transposase (IS6501 family)	y4oN	
y4wM	pu. ABC transporter binding prot.	y4tA	pu. transposase-associated ATP-binding protein; id. to y4bM & y4kI (sim. to IS21/IS1162 family)	y4pH	id. to y4bA
y4xM	<i>permease-type protein</i> ; sim. to <i>E. coli</i> YceE and to tetracycline transporters	y4tB	pu. transposase; id. to y4bL & y4kI (sim. to IS21/IS1162 family)	y4pI	id. to y4bB
protein export/secretion		y4tH	pu. transposase-associated ATP-binding protein (sim. to IS21/IS1162 family)	y4pJ	id. to y4bC
y4yI	<i>hrcN</i> pr. ATP synthase of a secretion system	y4uI	pu. transposase (sim. to IS21/IS1162 family)	y4pK	id. to y4bD
y4yK	<i>hrcQ</i> pr. secretion system protein (FliN/MopA/SpaO family)	y4uE*	pu. transposase (sim. to IS110 family)	y4rI	
y4yL	<i>hrcR</i> pr. secretion system protein (FliP/MopC/SpaP family)	y4zB*	pu. transposase (sim. to IS4 family)	y4rL*	
y4yM	<i>hrcS</i> pr. secretion system protein (FliQ/MopD/SpaQ family)	unknown function		y4rM*	
y4yN	<i>hrcT</i> pr. secretion system protein (FliR/MopE/SpaR family)	(with database homologue)		y4zA*	
y4yO	<i>hrcU</i> pr. secretion system protein (FliB/HrpN/SpaS family)	derived from IS element-like sequences			
y4yR	pr. secretion system protein (FliPEP family)	y4aQ	sim. to ORFs in <i>R. meliloti</i> & <i>A. tumefaciens</i> Ti plasmid	not linked to IS elements	
y4xJ	secretion system protein (PulD family)	y4hO	sim. to ORFC from <i>R. leguminosarum</i> symbiotic plasmid	y4aK	y4iH y4mD
chemotaxis		y4hP	sim. to ORFs in <i>R. meliloti</i> & <i>A. tumefaciens</i> Ti plasmid	y4aO	y4iI y4mG
y4iA	pr. chemoreceptor	y4hQ	sim. to ORF-3 in <i>A. rhizogenes</i> plasmid pRIA4B	y4aR	y4iJ y4mH
y4iS	pr. chemoreceptor	y4iC	sim. to ORFC from <i>R. leguminosarum</i> symbiotic plasmid	y4aS	y4iK* y4nF
electron transfer		y4jD	sim. to ORFs in <i>R. meliloti</i> & <i>A. tumefaciens</i> Ti plasmid	y4bG	y4iL* y4nI
y4iB	pu. 3Fe-3S ferredoxin	y4jI	sim. to ORFs in <i>R. meliloti</i> & <i>A. tumefaciens</i> Ti plasmid	y4bH	y4iM* y4nK
transcriptional regulation		y4kI	sim. to ORFs in <i>R. meliloti</i> & <i>A. tumefaciens</i> Ti plasmid	y4bN	y4iN* y4oB
y4aM	pu. DNA-binding protein according to <i>Rsp. rubrum</i> protein homologue	not linked to IS elements		y4bO	y4iF y4oC
y4aP	mucR Ros/MucR homologue	y4aN	sim. to <i>A. rhizogenes</i> pRIA4B replication region ORF3	y4bP	y4iG y4oD
y4cH	pr. cold shock regulator	y4bI	sim. to <i>H. influenzae</i> HI1631	y4bQ	y4iH y4oE
y4dJ	PbsX family H-T-H regulator	y4bK	sim. to <i>Ps. aeruginosa</i> PAH cluster ORF1	y4bR	y4iI y4oF
y4fK	AraC family H-T-H regulator	y4bM	sim. to <i>E. coli</i> HipA & <i>H. influenzae</i> HI0665	y4bS	y4iJ y4oG
y4mF	pr. H-T-H regulator; sim. to phage p22 C2	y4cO*	sim. to mitochondrial intron encoded ORFs & <i>E. coli</i> ORF319	y4bT	y4iK y4oH
y4mQ	LysR family H-T-H regulator	y4cP	sim. to <i>A. tumefaciens</i> Ti plasmid ORF2&3 in conjugal transfer region 1	y4bU	y4iL y4oI
y4pA	pr. sigma54-dependent H-T-H regulator	y4cQ	sim. to N-terminus of <i>E. coli</i> pRP4 & pR751	y4bV	y4iM y4oJ
y4pD	Ros/MucR homologue	y4cR	TraC	y4bW	y4iN y4oK
y4qH	LuxR family H-T-H regulator	y4fR	sim. to <i>Sh. flexneri</i> IpaH 7.8 & <i>Y. pestis</i> YopM	y4bX	y4iO y4oL
y4sM*	AsnC/Lrp family H-T-H regulator			y4bY	y4iP y4oM
y4tD	AsnC/Lrp family H-T-H regulator			y4bZ	y4iQ y4oN
y4wC	pu. DNA-binding protein according to <i>Rsp. rubrum</i> protein homologue			y4bA	y4iR y4oO
y4xI	Signal transduction-type regulator			y4bB	y4iS y4oP
integration/recombination				y4bC	y4iT y4oQ
y4bF	pu. transposase (sim. to IS1202)			y4bD	y4iU y4oR
y4bL	pu. transposase; id. to y4kI & y4tB (sim. to IS21/IS1162 family)			y4bE	y4iV y4oS
y4bM	pu. transposase-associated ATP-binding protein; id. to y4kI & y4tA (sim. to IS21/IS1162 family)			y4bF	y4iW y4oT
y4cG	pr. DNA invertase 'resolvase-type'			y4bG	y4iX y4oU
y4eF	pu. integrase/recombinase 'phage-type'			y4bH	y4iY y4oV
y4gC	pu. integrase/recombinase 'phage-type'			y4bI	y4iZ y4oW
y4hN	sim. to IS2/IS1312/IS866 proteins			y4bJ	y4jA y4oX

The nomenclature system for the ORFs is described in the Fig. 1 legend. General abbreviations: amino acids (aa), identical (id.), probable (pr.), putative (pu.), similar (sim.), hypothetical (hyp.), thiamine pyrophosphate (TPP), alanine (ala), insertion sequence (IS). Abbreviations of organisms: *Agrobacterium* (A.), *Azorhizobium* (Az.), *Azotobacter* (Azo.), *Bacillus* (B.), *Bradyrhizobium* (Br.), *Candida* (C.), *Desulfovibrio* (D.), *Escherichia* (E.), *Erwinia* (Erw.), *Haemophilus* (H.), *Mycobacterium* (M.), *Mycococcus* (Myx.), *Pseudomonas* (Ps.), *Rhizobium* (R.), *Rhodospirillum* (Rsp.), *Shigella* (Sh.), *Streptomyces* (St.), *Vibrio* (V.), *Zymomonas* (Z.). Asterisk indicates possibly fragmentous gene.

some essential *fix* loci are probably carried on the chromosome²³ (V. Viprey, personal communication).

One way of gaining insight into the function of genes is to follow their expression under different conditions. To do this, RNA was hybridized against filters containing amplified portions of 113 genes and gene fragments stretching from y4uA to y4bA. Under the conditions tested, transcripts were produced from most of the ORFs, but only a few genes are expressed in liquid medium (Fig. 5). Induction with daidzein rapidly increased transcript levels of y4vC, y4xL, y4xO, y4yB, y4yP, y4yQ (all of unknown function), y4xP (cysteine synthase) and y4zF (*syrM2*). Others, such as y4wF (a putative monooxygenase), y4wE (an aminotransferase), y4wM (an ABC-transporter binding protein), y4xI (a signal transduction-type regulator), y4xK (unknown), as well as *nolBTUV* and *hrcNQRST* are induced later. As expected, those genes involved in nitrogen fixation (both copies of *nifKDH*, and *nifE*), as well as y4wA, y4wb (which encodes zinc peptidases), y4wC (a DNA-binding protein), y4wD (a permease), y4aP (*mucR*) and y4aQ (unknown) are strongly expressed in nodules.

In addition to bacterial Nod factors, other signal molecules are probably necessary for the establishment of an effective symbiosis. This is exemplified by a rhamnose-rich extracellular polysaccharide that is involved in bacteroid development and nitrogen fixation in *Vigna*. A single locus encodes the complete biosynthetic pathway of dTDP-rhamnose from glucose-1-phosphate (y4gH, y4gF, y4gI, y4gH and *fixF*), while the y4gI gene product is probably needed for synthesis of rhamnose-rich lipopolysaccharides from dTDP-rhamnose²⁴.

Polypeptides and proteins are also probably involved in symbiosis. Six ORFs downstream of *nolXWBTUV* show strong homology to components of the type III secretion machinery of animal and plant pathogens. Homologues of these genes (*hrcN*, y4yI, *hrcQ*, *hrcR*, *hrcS*, *hrcT* and *hrcU*) are responsible for secretion of various proteins, and have been sequenced in the closely related bacterium *R. fredii* strain USDA257 (EMBL accession no. L12251). Secretion of

five genistein-induced proteins of USDA257 is dependent on a functional *nolXWBTUV* locus²⁵. Both in mammalian and plant pathogens, most of the proteins exported by the type III secretion machinery are pathogenicity determinants. Homologues of YopM of *Yersinia pestis* (y4fR), as well as avirulence genes of *Pseudomonas syringae* (y4fO) and *Xanthomonas campestris* (y4zC), were also found, reinforcing the idea that symbiotic and pathogenic interactions share common molecular mechanisms.

Surprisingly, mosaic sequences and insertion sequences comprise 18% of pNGR234a, and resemble those of diverse eubacteria (*Agrobacterium*, *Bacillus*, *Escherichia*, *Pseudomonas* and *Rhizobium*). Many are clustered between nucleotides 300,000 and 390,000 (Fig. 1). Other insertion and mosaic sequences divide the replicon into large blocks of functionally related genes (*oriV-oriT*, *nif*, *fix*, *hcn*; Fig. 4), suggesting that NGR234 has functioned as a 'transposon trap'. The G+C contents of these insertion and mosaic sequences are 3% higher than that of pNGR234a (58.5%) which, in turn, is 3.7 mol% less than the 62.2 mol% calculated for the entire genome²⁶. Several genes, especially those involved in Nod-factor and polysaccharide synthesis, have low G+C contents (45–55 mol%), raising the possibility that *nod* genes evolved separately from *nif* and *fix* genes (G+C content, 59 mol%). Although transposition of these insertion elements has not been demonstrated, transfer of plasmids amongst rhizobia in the legume rhizosphere²⁷ and to other non-symbiotic bacteria in fields²⁸ indicates that lateral transfer of genetic information has helped model symbiotic potential.

The high proportion of insertion and mosaic sequences and the well-conserved *Agrobacterium* conjugal transfer loci have broad implications for the evolution of symbioses between legumes and *Rhizobium*. Both genera are closely related²⁹, and the similarity of basic plasmid functions suggests a common origin. It is thus possible that an *Agrobacterium*-like progenitor gathered symbiotic genes through transposition with other soil bacteria, or that the progenitor was *Rhizobium* which assimilated opine and hormone genes from the plant³⁰. Transposable elements shape evolution in

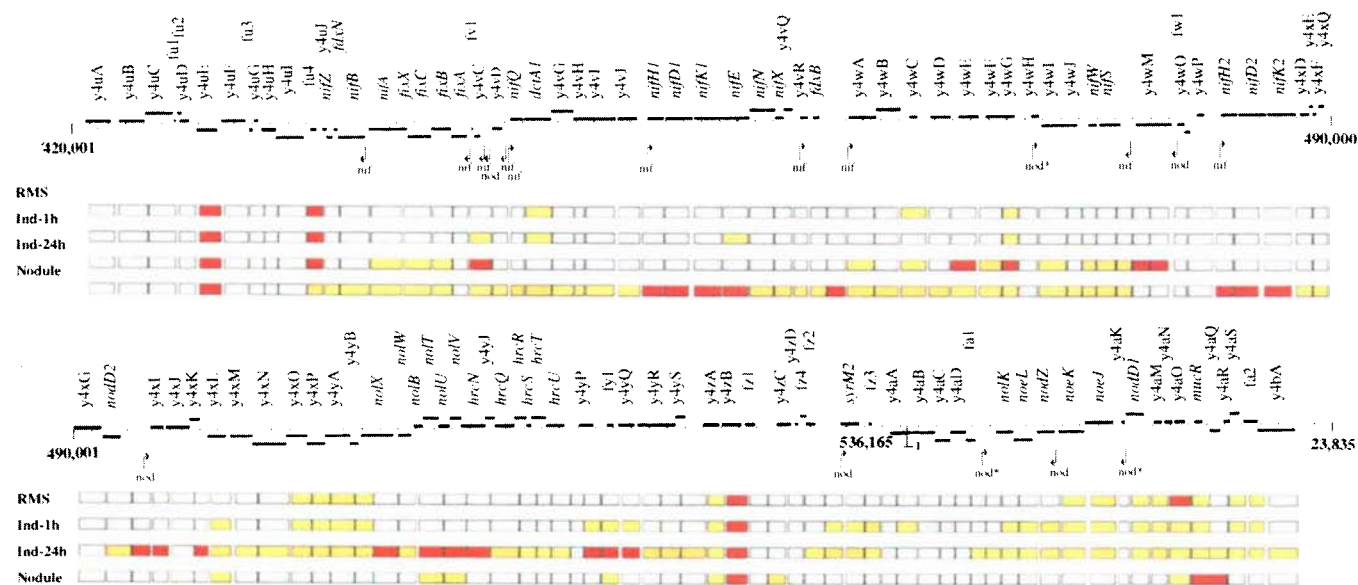


Figure 5 High-resolution transcription map of the 137-kb region encompassing the *nif* and *fix* clusters as well as the *hcn* locus. Predicted coding regions and gene fragments are shown on each strand (black rectangles). Positions of the corresponding amplified fragments are shown directly under the ORF map. Positions and orientations of the putative *nod*-box (*nod*) and NifA- σ^{54} (*nif*) promoter sequences are marked with arrows. Probable non-functional *nod* boxes are marked with asterisks. RNA probes were prepared from: RMS, NGR234 cells

grown at 28°C in liquid *Rhizobium* minimal supplemented with succinate (RMS)²³ medium; Ind-1h, cells grown in RMS followed by a 1-h induction with 2×10^{-7} M daidzein; Ind-24h, RMS-grown rhizobia collected 24 h after induction with daidzein; Nodule, bacteroids purified from Fix⁺ nodules of *V. unguiculata* inoculated with NGR234. Intensity of the hybridization signals detected on autoradiograms are colour coded as follows: yellow, weak; orange, medium; red, strong; and white, no signal.

different ways: they could 'pick up and carry' symbiotic or pathogenicity genes from one organism to another by forming composite transposons, or induce deletions, duplications, inversions and replicon fusions³¹. Unfortunately, there is no formal proof that the insertion and mosaic sequences of NGR234 have played these roles, although it seems likely, given the mosaic structure of the replicon. A striking example is the cytochrome P450 cluster (y4kS to y4kV, and y4lA to y4lD) which is 10% richer in G+C content than the rest of the plasmid, and has proteins with 83–93% similarity to those of *Bradyrhizobium*. Thus conjugal transfer and transposition have probably directed the evolution of bacteria which in one case learned genetically to colonize plants, and in the other to enter them. □

Methods

Sequencing and sequence analysis. Dye-terminator methods were used to 'shotgun' sequence a canonical array of cosmids⁷. Coding regions were detected using a mixture of intrinsic and extrinsic methods³². GeneMark predictions were based on matrices developed for *R. leguminosarum* and *R. meliloti*. BLAST and FASTA were used for similarity searches against Swiss-Prot, TrEMBL, GenBank and EMBL. Signal sequences, transmembrane segments and other characteristic domains were identified in putative proteins using the PC/Gen package. Definitions of protein families can be retrieved from the Prosite database (<http://www.expasy.ch/sprot/prosite.html>)¹². Scans for homologies were last performed in November 1996 using resources at NCBI (<http://ncbi.nlm.nih.gov>), ExPASy (<http://expasy.ch>) and EBI (www.ebi.ac.uk). Searches for consensus promoter sequences were performed using FINDPATTERNS (WASP version 8, Genetics Computer Group, Madison, Wisconsin).

Transcription analysis. We made 113 PCR products ranging from ~600 bp to ~2,000 bp between y4uA (5'-end, bp 420,774) and y4bA (3'-end, bp 21,758) using cosmid DNA as the template and primer pairs designed to amplify the discrete ORFs or intergenic regions. In some cases, M13 phage DNA (from the sequencing libraries) was used to produce PCR products. Portions of the amplified fragments were then separated on 0.8% agarose gels, and vacuum-blotted onto GeneScreen Plus Nylon membranes (NEN).

RNA extraction, labelling and hybridizations. Cell cultures at an absorbance at 600 nm of 0.4–0.5 were collected by centrifugation. Bacteroids were isolated from nodules crushed in liquid nitrogen and resuspended in sterile water. Debris was removed by filtration and bacteroids were recovered by centrifugation (4,000g, 5 min). Purification, radioactive labelling of RNA from bacteroids and collected cells, and hybridization conditions were as described¹⁸. No unlabelled competitor RNA was added to the prehybridization solution. Filters were exposed for 6–72 h, and the hybridization signals were grouped according to intensity: no signal, weak, medium and strong.

Received 14 February; accepted 14 April 1997.

- Fellay, R., Rochepeau, P., Relic, B. & Broughton, W. J. in *Pathogenesis and Host Specificity in Plant Diseases. Histopathological, Biochemical, Genetic and Molecular Bases* (eds Singh, U. S., Singh, R. P. & Kohmoto, K.) 199–220 (Pergamon, Oxford, 1995).
- Spaink, H. P. Regulation of plant morphogenesis by lipo-chitin oligosaccharides. *Crit. Rev. Plant Sci.* **15**, 559–582 (1996).
- Dénarié, J., Debellé, F. & Promé, J.-C. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signalling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* **65**, 503–535 (1996).
- Fischer, H.-M. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* **58**, 352–386 (1994).
- Lewin, A. et al. Multiple host-specificity loci of the broad host-range *Rhizobium* sp. NGR234 selected using the widely compatible legume *Vigna unguiculata*. *Plant Mol. Biol.* **8**, 447–459 (1987).
- Broughton, W. J., Heycke, N., Meyer, Z. A., H. & Pankhurst, C. E. Plasmid-linked *nif* and "nod" genes in fast-growing rhizobia that nodulate *Glycine max*, *Psophocarpus tetragonolobus*, and *Vigna unguiculata*. *Proc. Natl Acad. Sci. USA* **81**, 3093–3097 (1984).
- Freiberg, C., Perret, X., Broughton, W. J. & Rosenthal, A. Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable "Sequenase": a beginning. *Genome Res.* **6**, 590–600 (1996).
- Perret, X., Broughton, W. J. & Brenner, S. Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234. *Proc. Natl Acad. Sci. USA* **88**, 1923–1927 (1991).
- Fraser, C. M. et al. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
- Fleischmann, R. D. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
- Morrison, N. A. et al. Mobilization of a *sym* plasmid from a fast-growing cowpea *Rhizobium* strain. *J. Bacteriol.* **160**, 483–487 (1984).
- Bairoch, A., Bucher, P. & Hofmann, K. The PROSITE database, its status in 1995. *Nucleic Acids Res.* **24**, 189–196 (1996).
- Zhang, L., Murphy, P. J., Kerr, A. & Tate, M. E. *Agrobacterium* conjugation and gene regulation by *N*-acyl-L-homoserine lactones. *Nature* **362**, 446–448 (1993).
- Piper, K. R., Beck von Bodman, S. & Farrand, S. K. Conjugation factor of *Agrobacterium tumefaciens* regulates Ti plasmid transfer by autoinduction. *Nature* **362**, 448–450 (1993).

- Price, N. P. J. et al. Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated and fucosylated, nodulation signals that are O-acetylated or sulphated. *Mol. Microbiol.* **6**, 3575–3584 (1992).
- Leigh, J. A. & Walker, G. C. Exopolysaccharides of *Rhizobium*: synthesis, regulation and symbiotic function. *Trends Genet.* **10**, 63–67 (1994).
- Hanin, M. et al. Sulphation of *Rhizobium* sp. NGR234 Nod factors is dependent on *noeE*, a new host specificity gene. *Mol. Microbiol.* (in the press).
- Fellay, R., Perret, X., Viprey, V., Broughton, W. J. & Brenner, S. Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol. Microbiol.* **16**, 657–667 (1995).
- Jabbouri, S. et al. Involvement of *nodS* in *N*-methylation and *nodU* in 6-O-carbamoylation of *Rhizobium* sp. NGR234 Nod factors. *J. Biol. Chem.* **270**, 22968–22973 (1995).
- van Slooten, J. C., Cervantes, E., Broughton, W. J., Wong, C. H. & Stanley, J. Sequence and analysis of the *rpoN* sigma factor gene of *Rhizobium* sp. strain NGR234, a primary coregulator of symbiosis. *J. Bacteriol.* **172**, 5563–5574 (1990).
- Badenoch-Jones, J., Holton, T. A., Morrison, C. M., Scott, K. F. & Shine, J. Structural and functional analysis of nitrogenase genes from the broad-host range *Rhizobium* strain ANU240. *Gene* **77**, 141–153 (1989).
- van Slooten, J. C., Bhuvanavari, T. V., Bardin, S. & Stanley, J. Two C4-dicarboxylate transport systems in *Rhizobium* sp. NGR234: rhizobial dicarboxylate transport is essential for nitrogen fixation in tropical legume symbiosis. *Mol. Plant Microbe Interact.* **5**, 179–186 (1992).
- Broughton, W. J. et al. Identification of *Rhizobium* plasmid sequences involved in recognition of *Psophocarpus*, *Vigna*, and other legumes. *J. Cell Biol.* **102**, 1173–1183 (1986).
- Jabbouri, S. et al. in *The Biology of Plant-Microbe Interactions* (eds Stacey, G., Mullin, B. & Gresshoff, P.) 319–324 (International Society of Molecular Plant-Microbe Interactions, St Paul, MN, 1996).
- Krishnan, H. B., Kuo, C.-I. & Pueppke, S. G. Elaboration of flavonoid-induced proteins by the nitrogen-fixing soybean symbiont *Rhizobium fredii* is regulated by both *nodD1* and *nodD2*, and is dependent on the cultivar-specificity locus, *noXWBTUV*. *Microbiology* **141**, 2245–2251 (1995).
- Broughton, W. J., Dilworth, M. J. & Passmore, I. K. Base ratio determination using unpurified DNA. *Anal. Biochem.* **46**, 164–172 (1972).
- Broughton, W. J., Samrey, U. & Stanley, J. Ecological genetics of *Rhizobium meliloti*: symbiotic plasmid transfer in the *Medicago sativa* rhizosphere. *FEMS Microbiol.* **40**, 251–255 (1987).
- Sullivan, J. T., Patrick, H. N., Lowther, W. L., Scott, D. B. & Ronson, C. W. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc. Natl Acad. Sci. USA* **92**, 8985–8989 (1995).
- Martinez-Romero, E. & Cabellero-Mellado, J. *Rhizobium* phylogenies and bacterial genetic diversity. *Crit. Rev. Plant Sci.* **15**, 113–140 (1996).
- Tepfer, D. in *Plant Microbe Interactions* (eds Kosuge, T. & Nester, E.) 294–342 (McGraw Hill, New York, 1989).
- Galas, D. J. & Chandler, M. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 109–162 (American Society of Microbiology, Washington DC, 1989).
- Borodovsky, M., Rudd, K. E. & Koonin, E. V. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.* **22**, 4756–4767 (1994).

Acknowledgements. We thank E. Michaelis, D. Schnabelrauch, S. Landmann, S. Förste, K. Blechschmidt and G. Nordesiek for technical assistance; B. Drescher and D. Bauer for computational assistance; M. Y. Borodovsky and W. S. Hayes for GeneMark matrices; O. White for helping to create Fig. 1; S. Farrand and P. Murphy for advice on the conjugal transfer experiments; D. Gerber and S. Relic for help with many aspects of this work; and P. Rochepeau and V. Viprey for making their unpublished data available.

Correspondence and requests for materials should be addressed to A.R. (e-mail: arosenth@imb-jena.de). The sequence of pNGR234a has been deposited in EMBL/GenBank (accession no. U00090). Putative proteins are annotated in Swiss-Prot. Further information can be retrieved from the World-Wide Web (<http://genome.imb-jena.de/archives.html>).

Task difficulty and the specificity of perceptual learning

Merav Ahissar*† & Shaul Hochstein‡

* Center for Higher Brain Functions, Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

‡ Center for Neural Computation, Department of Neurobiology, Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel

Practising simple visual tasks leads to a dramatic improvement in performing them. This learning is specific to the stimuli used for training. We show here that the degree of specificity depends on the difficulty of the training conditions. We find that the pattern of specificities maps onto the pattern of receptive field selectivities along the visual pathway. With easy conditions, learning generalizes across orientation and retinal position, matching the spatial generalization of higher visual areas. As task difficulty increases, learning becomes more specific with respect to both orientation and position, matching the fine spatial retinotopy exhibited by lower areas. Consequently, we enjoy the benefits of

† Present address: Keck Center for Integrative Neuroscience, UCSF, San Francisco, California 94143-0732, USA.

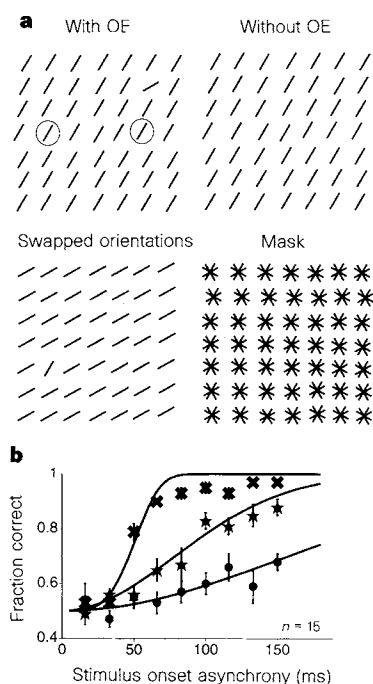


Figure 1 Odd-element detection scheme. **a**, The stimulus array with and without an odd element (OE). Circles (not shown on screen) indicate target positions under the 2-position paradigm. **b**, Average performance improvement (all positions paradigm) from the first and second thirds of the initial session (circles, stars) to the post-training final session (crosses). Lines are Quick psychometric functions fitted to the data. Training induces a leftward shift and steepening of the psychometric curve, substantially decreasing the threshold (SOA for 82% correct). Threshold is the performance measure used in Fig. 2a.

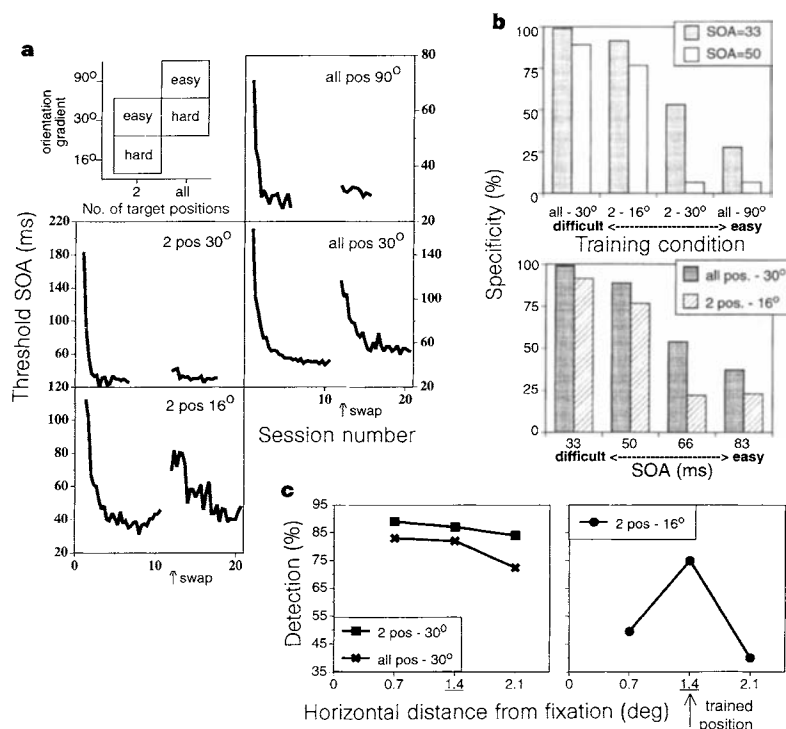


Figure 2 Linkage between difficulty and learning specificity. Comparison of learning specificity under 4 training schemes: target could appear at any array position (with equal probabilities) except fixation (all: right column) or at one of 2 positions (2 pos: left column). Target orientation deviated from that of distractors by 90°, 30° or 16° (top, middle and bottom rows, respectively). Difficult conditions are 2 pos-16° or all-30° (see Fig. 1a) and easy conditions are 2 pos-30° or all-90° (inset). Transfer to new orientations and positions was larger for easy than for difficult task conditions: **a**, Learning curves and orientation specificity for four training conditions. Average threshold as a function of (thirds of) session number for each of the 4 subject groups. Note improvement, and transfer for easy conditions. (Bottom plot begins at the second-third session, as subjects barely noticed the target initially; threshold was 220 ms.) **b**, Comparison of relative specificity (performance decrement following orientation swapping as per cent of total gained improvement) for 4 conditions and different SOAs: top, specificity

(for fixed SOA) decreases with decreased task difficulty; bottom, specificity (for fixed paradigm) decreases with increased SOA. **c**, Position specificity of learning 2 positions (left/right of fixation; at 1.4° eccentricity). Detection at short SOA (33 ms) is plotted (for trained and neighbouring positions) as a function of horizontal distance from fixation. Left, training easy 2 pos-30° yields improvement along the horizontal meridian, including interpolated and extrapolated positions. Detection measured during session testing all positions following 2-position training (squares). For comparison, asymptotic detection for the group trained on all pos-30° is also shown (crosses). Training with 2 positions leads to superior performance at the trained positions and generalizes to the (horizontally) neighbouring positions. Right, training difficult 2 pos-16° leads to more spatially specific improvement. Testing 3 subjects with 2 proximal or distal positions results in poorer detection than at the trained positions. Repeated tests with trained position showed s.d. <2% for each of the subjects.

learning generalization when possible, and of fine grain but specific training when necessary. The dynamics of learning show a corresponding feature. Improvement begins with easy cases (when the subject is allowed long processing times) and only subsequently proceeds to harder cases. This learning cascade implies that easy conditions guide the learning of hard ones. Taken together, the specificity and dynamics suggest that learning proceeds as a countercurrent along the cortical hierarchy. Improvement begins at higher generalizing levels, which, in turn, direct harder-condition learning to the subdomain of their lower-level inputs. As predicted by this reverse hierarchy model, learning can be effective using only difficult trials, but on condition that learning onset has previously been enabled. A single prolonged presentation suffices to initiate learning. We call this single-encounter enabling effect 'eureka'.

The intricate processing of visual stimuli by the brain is far from being understood. However, two principles of neuronal representation are generally accepted. (1) The sequence of processing is related to an anatomical–physiological hierarchy of visual areas with multiple forward, feedback and cross-area pathways^{1,2}. (2) Neuronal responses along the hierarchy begin with selective tuning to basic spatial dimensions (for example, retinal location, stimulus orientation and size). They gradually generalize over spatial dimensions and become increasingly specific to complex dimensions of the stimulus³.

The ability of adults to improve in almost every simple perceptual task is remarkable. Psychophysical studies have characterized the specificity of this improvement: when training with a consistent, limited range of stimuli, improvement is specific to spatial stimulus attributes such as retinal position^{3–8}, orientation^{4–13}, size^{10,11,13} and, in some cases, even the trained eye⁴. These findings indicate an early cortical site of plasticity where receptive fields retain fine selectivity along basic spatial dimensions. However, no less prominent is the huge variability of specificity within and across studies, spanning the full range between complete specificity and none (for example, for orientation^{5,12–14}, position^{6,7,15}, motion direction^{8,16,17} or eye^{4,15,18}). Characterizing learning as early cortical site modification is not compatible with this large variability.

We now find a consistent pattern underlying learning specificity across tasks and conditions: the more difficult the behavioural situation, the more specific the practice effects. On the basis of this rule, we propose a cortical learning scheme in which training under different conditions affects different cortical sites and processes¹⁷. In particular, learning proceeds in a reverse hierarchy, in the direction countercurrent to bottom-up stimulus processing, following instead the path of top-down attention effects¹¹.

The paradigm we used is simple feature detection^{19,20}. Subjects reported the presence or absence of an oddly oriented element in a light-bar array (Fig. 1a). As target orientation deviated greatly from that of distractors, detection was easy for prolonged stimulus presentations. Task difficulty stemmed from brief presentation followed by a masking stimulus after a variable time interval (the stimulus-to-mask onset asynchrony, SOA). At short SOAs, performance is at chance level, whereas at long SOAs complete detection is achieved (for example, the psychometric curves in Fig. 1b). Practice induces dramatic improvement until a steady performance level is reached (Figs 1b and 2a). We then test and train subjects with manipulated stimuli to measure learning specificity.

Figure 2 compares learning and specificity characteristics under four training conditions. The right column of Fig. 2a illustrates orientation specificity with the target at any array position for large (90°, top) and intermediate (30°, middle right) target-distractor orientation differences (or gradients). Under the easy 90° gradient, learning was general and almost complete transfer was found when swapping target-distractor orientations. On the other hand, for the 30° gradient learning was specific, and improved performance with swapped orientations required relearning. Similarly, for the target at one of two horizontal positions, there was no specificity for the

easier 30° gradient and substantial orientation specificity for the difficult 16° gradient (left column). Increasing positional difficulty (uncertainty and mean eccentricity) also increased specificity (compare Fig. 2a, middle row, left: 2 positions; right: all positions). Consistently, stimulus specificity increased with increased task difficulty induced by increased target eccentricity and positional uncertainty, or by decreased target-distractor orientation gradient (Fig. 2b, top). The dependence of learning specificity on discrimination difficulty implies that different learning processes are activated under these different stimulus conditions. In particular, improvement for difficult orientation discriminations depends on modification of neurons that are finely tuned to orientation, whereas modifications in easy conditions involve processes that do not retain fine orientation separability.

We tested whether difficulty in orientation discrimination also leads to position specificity. Figure 2c shows position specificity following training with two target positions, compared with a control group who learned with all positions. The left plot shows that (testing with target at any array position, following two-position training), improvement for the 30° gradient generalized to horizontal locations between and beyond those actually activated by target presence. For the difficult 16° gradient, position generalization was significantly reduced, so that after training, detection at the trained position substantially exceeded that at the proximal and distal positions (Fig. 2c, right). Thus, increasing orientation difficulty induces both orientation and position specificity. Increasing positional difficulty also induced orientation specificity (Fig. 2a, middle row, compare right and left). Thus, there is bidirectional coupling of orientation and position.

We investigated whether specificity would also be affected when increasing difficulty by decreasing available processing time (shortening SOA) and thereby decreasing signal discriminability. Figure 2b (bottom) shows specificity as a function of SOA. For both (difficult) training conditions, specificity decreases with increasing SOA. Specificity varies with SOA for identical spatial stimuli within the same subjects. This dependence suggests that, even within a given paradigm, and for the same subject, different learning processes are activated by blocks of different SOAs. The dependence on SOA follows the same rule as the dependence on difficulty stemming from orientation gradient or target position: increasingly difficult conditions (short SOAs) activate increasingly specific learning processes.

This pattern of transfer across stimuli for easy condition training versus specific learning for difficult conditions is consistent with findings in motion discrimination¹⁷ and reconciles other seemingly inconsistent data which have been found for various behavioural tasks and testing paradigms. Learning to discriminate motion direction was found to be highly specific to the trained direction, when training a difficult direction discrimination (3° gradient)^{16,17}, but there was substantial transfer to other motion directions when training an easy discrimination¹⁷. Learning difficult conjunction detection is position-specific, whereas learning easy feature detection is not⁷. Learning texture discrimination—a task that is similar to feature detection—was found to be position-specific when conditions were difficult⁴. In texture discrimination, the initial learning of easy cases (long SOAs) transferred across eyes, whereas the more difficult later learning phase did not²¹. Improvement in orientation discrimination was more orientation-specific under difficult peripheral training conditions⁵ than under central easier training¹⁴. Finally, complex pattern recognition was more location-specific for more difficult discriminations^{22,23} and its learning could be largely position-specific⁸.

Can learning of difficult cases (at low representation levels) occur right from the start? If so, improvement at the various (blocks of) SOAs would proceed in parallel. However, analysing learning curves for each SOA separately shows different dynamics (Fig. 3, left). Learning at each SOA is sigmoidal: the rising phase starts earliest for

the easiest SOAs, and only after substantial improvement is obtained for these, does improvement begin for harder, shorter SOAs. Improvement cascades gradually from easy to harder conditions, even though different SOA blocks are intermixed. At shorter SOAs, improvement is slower, reaching asymptote after more than one or two sessions (Fig. 3, middle). Subsequent orientation swapping disrupts the improvement gained for difficult short SOAs more than it disrupts that gained for the long easy ones (Fig. 3, right, and Fig. 2b, bottom).

Taken together, the pattern of specificities and the learning cascade indicate that learning proceeds from higher to lower areas, in the countercurrent direction. The easy-to-difficult learning cascade further suggests that higher-level improvement may be necessary to allow low-level learning. We conclude that reverse hierarchy attentional search²⁴⁻²⁷ is applied for determining the site of task performance and learning plasticity (Fig. 4).

The reverse-hierarchy model predicts that if training is limited to difficult stimuli, improvement will not occur for lack of a pointer to the appropriate lower level site of detailed spatial representation. It had been found that introducing easy discriminations in a subjective contour task dramatically facilitated abrupt learning of hard cases²⁸. We tested this prediction in the context of popout detection²⁹. A group of subjects ($n = 23$) was trained for a whole session with only 50 ms SOA (without feedback). This SOA was chosen as most subjects who started out with chance-level performance

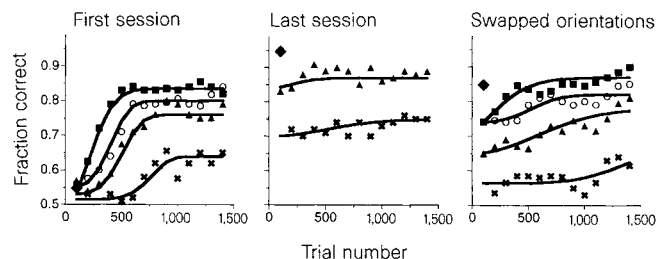


Figure 3 Dynamics of improvement for each SOA. Average fraction correct is plotted as a function of trial number for the group practicing all-30°. Short (20 trial) blocks of different SOAs were presented in pseudorandom order (so that, on average, various SOAs were presented within each 100-trial bin plotted here). During the first session (left), performance for each SOA remained at first at near chance (50%), and improvement began only following some practice. The extent of this initial plateau was SOA-dependent, being short (several dozen trials) for long SOAs and long for short SOAs (~400 trials for 50 ms (triangles) SOA; >500 trials for 33 ms (crosses) SOA). Diamonds, 166 ms SOA; squares, 83 ms; circles, 66 ms. The slope of improvement was also SOA-dependent, being steeper for longer SOAs. Towards the end of the first session, performance stabilized. Further improvement was gained on subsequent sessions, as seen by comparison with asymptotic performance (centre panel). On the first session with swapped orientations (right), initial performance for long SOAs was 85% correct compared with 55% before initial-orientations learning (left). Thus, there was considerable transfer for easy SOAs. For shorter SOAs, a smaller fraction of the learning effect was transferred, and for 33-ms SOA, practically no transfer is seen. Moreover, subsequent improvement starts after a prolonged delay. At 50 ms SOA, initial and second learning profiles differ greatly: the original sigmoidal curve is replaced by a shallow curve; performance starts out better; improvement is markedly retarded. Points were averaged triangularly with neighbouring points.

managed to improve substantially when these were intermixed with easier blocks of trials (Fig. 3, left). As predicted, when only hard cases were trained, for most subjects no improvement was apparent throughout the session (Fig. 5a, left).

What is the minimum needed to initiate difficult trial learning? According to the model, perceiving one easy case can highlight the appropriate low-level region. To test minimal experience, we tested another group of subjects with the difficult 50-ms SOA procedure. This group was first shown a long (30 s) presentation of the whole array on the monitor screen, once with and once without an odd element. They were then tested immediately afterwards. Their results were categorically different: dramatic improvement was shown by most of this group (Fig. 5a, right). We term this large single-presentation learning-enabling the 'eureka' effect²⁹. Consistent results have been found in other paradigms involving, as already mentioned subjective contours²⁸ and texture discrimination³⁰.

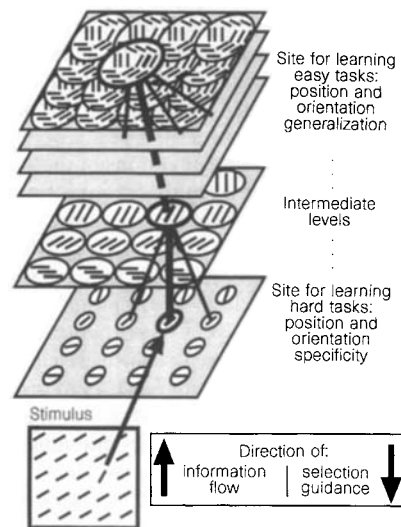


Figure 4 the reverse-hierarchy model. each circle denotes a neuronal group (for example, cortical column) and the line(s) inside it denote(s) its orientation preference. Enhanced lines mark bottom-up stimulus-activated neuron groups and paths interconnecting them. Stimulus is initially encoded by orientation-selective neurons, at the lowest level. For example, one group is shown encoding the target element bar. At subsequent levels, there is substantial convergence across position (second level) and orientation (top level). Thus, the orientation of the original target signal is gradually undiscriminated, and its salience diminished when neurons integrate over many orientations (and convergence refines more abstract features). Spatial attention and learning are initially directed at the highest, spatially generalizing levels. For easy conditions (such as large orientation gradient and long SOA), high levels suffice: even diminished salience leaves some distinction between target presence/absence. In this case, learning occurs at a high level (for example, by selecting among its inputs). If insufficient signal remains at high levels, subjects may be stuck. Alternatively, if high-level mechanisms are pinpointed by easy condition training, they may be used to direct attention and learning to appropriate lower-level mechanisms, within the subdomain of their inputs. The learning cascade (Fig. 3) may reflect reverse order search along the hierarchy. For these lower-level mechanisms, learning will be restricted to trained orientations and positions, reflecting increased spatial selectivity at this level. When a new stimulus is presented (as with swapped orientations), common stimulus characteristics suffice for easy (long SOA) performance; high-level mechanisms of the hierarchy, common for both orientation sets, result in some transfer (Fig. 2). For difficult cases (small orientation gradient and/or short SOA), the lower-level mechanisms by which learning effects occur are spatially selective and inappropriate for new task conditions. Countercurrent paths from the common top-level mechanism will have to be used again to redirect learning to different low-level sites. This redirection may account for short-SOA retarded learning when swapping orientations (Fig. 3, right).

The main prediction of the model is that easy-condition, high-level effects are generalized, whereas difficult condition, low-level learning is spatially specific. The eureka phenomenon depends on two stages: first, an easy (long) presentation enables learning; then, learning is effected during a session of difficult (brief SOA) trials. The model therefore predicts that eureka presentation should enable learning of a wider range of spatial stimuli than those actually learned during the subsequent hard cases training period. This was confirmed as follows: we divided the subjects exposed to prolonged presentation into 2 subgroups. One was enabled and tested with stimuli of unchanged target and distractor orientations, while the other was tested with swapped orientations. Comparing their initial learning curves (Fig. 5b) shows little difference in performance and learning rate. To a second session, we invited the subjects who had learned with test stimuli identical to eureka enabling stimuli. Now they were retested with swapped orientations (without a second enabling presentation). All had to relearn the task, and their second learning curves were very similar to their first ones with original orientations (Fig. 5c). We conclude that (high-level) eureka enabling transfers across spatial attributes, whereas (low-level) learning is stimulus-orientation specific.

Eureka presentation only enables learning, and does not further affect the learning process itself. Analysing the time course of

learning following eureka presentation, we found that improvement almost always began within 100 trials. In comparison, subjects who managed to learn without eureka presentation show large onset-time variability. (In contrast, duration of improvement was similarly distributed in the two groups.) As predicted by the reverse-hierarchy model, eureka presentation guides access to appropriate representations, but does not affect the learning process itself.

Thus, a broad body of evidence supports the rule that the harder the spatial discriminations trained, the greater the induced specificity with respect to basic spatial dimensions. We interpret this consistent trend as reflecting modifications at various levels of the visual system, in reverse hierarchical order. Introducing this concept reconciles many cases of apparently contradictory data. It does so by attributing apparently different strategies for improvement to modifications within different underlying neuronal sites. The selection of learning site follows a basic rule of centripetal, attention-guided search along the hierarchical levels of visual representation. Finally, the reverse hierarchy model has a variety of testable predictions, including the eureka effect and its characteristics. □

Methods

The stimulus was a 7×7 array of light bars with or without an odd element, with equal probabilities. Each stimulus element subtended $22 \times 1'$; distance

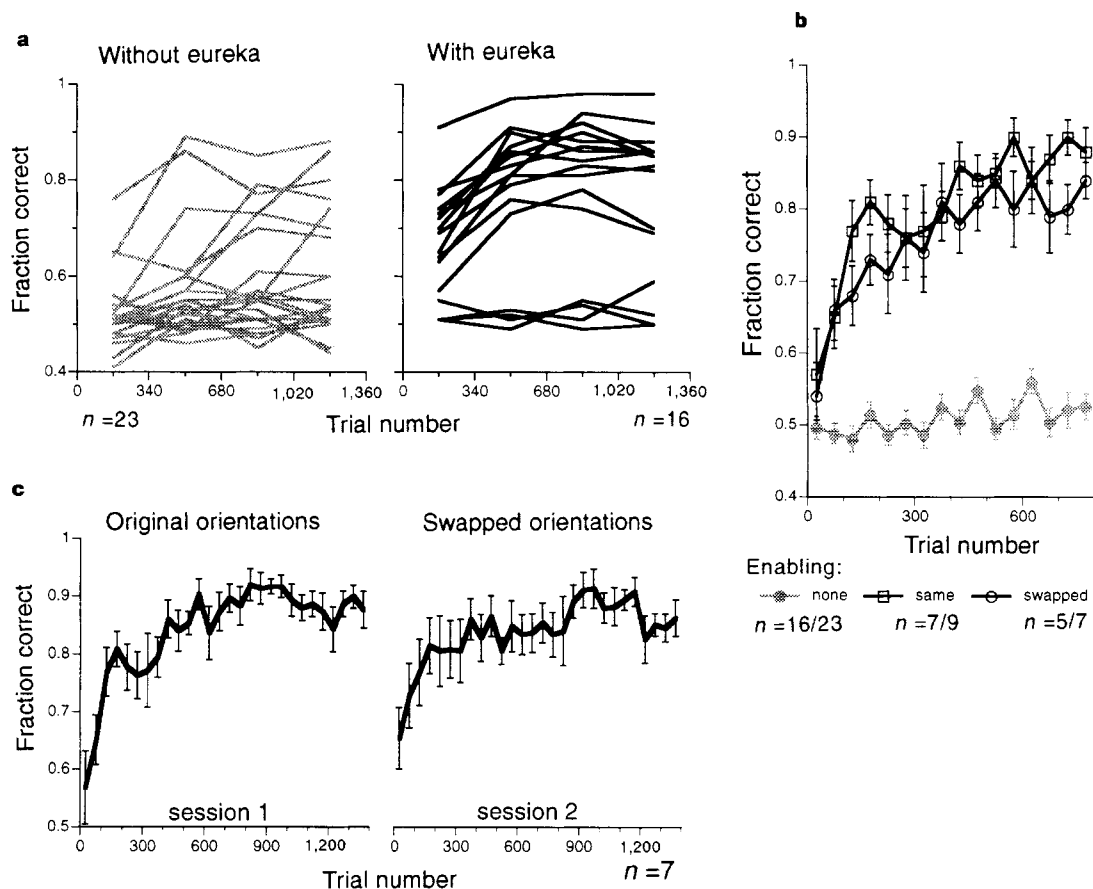


Figure 5 a, The 'eureka' effect. Most subjects fail to improve at all during their first session if they are presented with only threshold-level trials of 50 ms SOA (left: 16/23 failed to reach 60% correct even in the last quarter of the first session; each line shows the performance of 1 subject). The task was understood, as subjects received verbal instructions and were shown pen-and-pencil drawings of the array with and without an odd element. If subjects are allowed an extended view of the stimulus, once with and once without the odd element, most improve dramatically during the first session (right: 12/16 attained >70% correct). **b**, Transfer of the eureka enabling effect. Learning is similarly enabled when the eureka presentation is with the same stimulus as the test (squares; $n = 9$), or with

swapped orientations (circles; $n = 7$). However, showing a paper drawing sketch of the stimulus (disks) is ineffective. Curves and error bars show across-subject performance mean and s.e. (In the second half session, the swapped orientations group showed some performance decrement, perhaps due to increased fatigue.) **c**, Following enabling and training with one set of orientations (left), performance can be improved for the swapped orientations without additional enabling (right), though a similar period of training is needed with these. Thus (easy, high-level) eureka enabling transfers across orientation swapping but (difficult, low-level) learning is specific.

between element centres was 42.6'; circles in Fig. 1 (not shown on screen) indicate target positions under the 2-position paradigm. A mask followed each stimulus. Following training, target and distractor element orientations were swapped to test learning specificity. Trial temporal sequence was as follows: subjects pressed the 'ready' key in response to a fixation cross. Stimulus was followed by a mask after a variable SOA. Subjects then pressed a response key to indicate odd-element presence or absence. A pleasant computer tone rewarded correct responses. Stimuli were presented in blocks of 20 trials with the same SOA, 70 blocks per session (1,400 trials). Sessions began with 9 blocks starting from longest SOAs (183 or 150 ms) and gradually reaching shortest SOA (16 ms) in an interleaved manner. Based on performance in these initial blocks, the range of SOAs to be presented next was chosen so that the shortest SOA would be the longest in which the subject still performed near chance level (55% correct) and the longest SOA would be the shortest where the subject already showed near perfect performance (95% correct). Within that range, blocks were presented in pseudorandom sequence. The next range of SOAs was chosen on the basis of performance in these blocks, following the above criteria. Thus, performance was always kept ~75% correct^{11,13}. For the eureka experiments (Fig. 5), SOA was kept constant at 50 ms and a feedback tone was not given for correct responses.

Received 23 October 1996; accepted 27 February 1997.

1. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in primate visual cortex. *Cerebr. Cort.* **1**, 1–47 (1991).
2. Desimone, R. & Ungerleider, L. G. In *Handbook of Neuropsychology* (eds Boller, F. & Grafman, J.) Vol. 2 267–299 (Elsevier Science, Amsterdam, 1989).
3. Berardi, N. & Fiorentini, A. Interhemispheric transfer of visual information in humans: spatial characteristics. *J. Physiol.* **384**, 633–647 (1987).
4. Karni, A. & Sagi, D. Where practice makes perfect: Evidence for primary visual cortex plasticity. *Proc. Natl Acad. Sci. USA* **88**, 4966–4970 (1991).
5. Schoups, A. A., Vogels, R. & Orban, G. Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *J. Physiol.* **483**, 797–810 (1995).
6. Poggio, T., Fahle, M. & Edelman, S. Fast perceptual learning in visual hyperacuity. *Science* **256**, 1018–1021 (1992).
7. Treisman, A., Vieira, A. & Hayes, A. Automaticity and preattentive processing. *Am. J. Psychol.* **105**, 341–362 (1992).
8. Nazir, T. & O'Regan, J. K. Some results on translation invariance in the human visual system. *Spatial Vision* **5**, 81–100 (1990).
9. Ramachandran, V. S. & Braddick, O. Orientation specific learning in stereopsis. *Perception* **2**, 371–376 (1973).
10. Fiorentini, A. & Berardi, N. Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Res.* **21**, 1149–1158 (1981).
11. Ahissar, M. & Hochstein, S. The role of attention in early perceptual learning. *Proc. Natl Acad. Sci. USA* **90**, 5718–5722 (1993).
12. Fahle, M. & Edelman, S. Long term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Res.* **33**, 397–412 (1993).
13. Ahissar, M. & Hochstein, S. Learning pop-out detection: specificities to stimulus characteristics. *Vision Res.* **36**, 3487–3500 (1996).
14. Vogels, R. & Orban, G. A. The effect of practice on the oblique effect in line orientation judgments. *Vision Res.* **25**, 1679–1687 (1985).
15. Beard, B. L., Levi, D. M. & Reich, L. N. Perceptual learning in parafoveal vision. *Vision Res.* **35**, 1679–1690 (1995).
16. Ball, K. & Sekuler, R. A specific and enduring improvement in visual motion discrimination. *Science* **218**, 697–698 (1982).
17. Liu, Z. Learning a visual skill that generalizes. *NEC Tech. Rep.* 95–177 (NEC, Princeton, 1995).
18. Schoups, A. & Orban, G. A. Interocular transfer in perceptual learning of a pop-out discrimination task. *Proc. Natl Acad. Sci. USA* **93**, 7358–7362 (1996).
19. Treisman, A. & Gelade, G. A feature integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980).
20. Julesz, J. B. Textons, the elements of texture perception and their interactions. *Nature* **290**, 91–97 (1981).
21. Karni, A. & Sagi, D. The time course of learning a visual skill. *Nature* **365**, 250–252 (1993).
22. Dill, M. & Fahle, M. Limited translation invariance of human visual pattern recognition. *Percept. Psychophys.* (in the press).
23. Miller, J. Components of location probability effect in visual search tasks. *Exp. Psychol.: Hum. Percept. Perform.* **14**, 453–471 (1988).
24. Treisman, A. The binding problem. *Cur. Opin. Neurobiol.* **6**, 171–178 (1996).
25. Ullman, S. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex* **5**, 1–11 (1995).
26. Nakayama, K. In *Vision: coding and Efficiency* (ed. Blakemore, C.) 411–442 (Cambridge Univ. Press, Cambridge, 1991).
27. Tsotsos, J. K. Analyzing vision at the complexity level. *Behav. Brain Sci.* **13**, 423–469 (1990).
28. Rubin, N., Shapley, R. M., Nakayama, K. & Grosssetete, A. Abrupt learning in illusory contour perception. *Invest. Ophthalmol. Vis. Sci.* **37**, 859 (1996).
29. Ahissar, M. & Hochstein, S. Eureka: One shot viewing enables perceptual learning. *Invest. Ophthalmol. Vis. Sci.* **37**, 3183 (1996).
30. Papathomas, T. V., Gorea, A. & Feher, A. Attending to attributes in double conjunction texture segregation: the role of color, luminance and orientation. *Invest. Ophthalmol. Vis. Sci.* **37**, 2415 (1996).

Acknowledgements. We thank N. Rubin and R. Shapley for pointing out to us the importance of abrupt learning. We thank them and A. Treisman, M. Merzenich, H. Hock, S. Ullman, T. Flash, R. Malach, E. Ahissar and E. Zohary for suggestions and for discussion, and R. Laiwand and G. Kosminsky for running many of the experiments. This study was supported by grants from the US-Israel Binational Science Foundation and the Israel Science Foundation of the Israel Academy of Sciences and Humanities. M.A. was supported by a Feinberg fellowship at the Weizmann Institute of Science.

Correspondence should be addressed to S.H. (e-mail: shaul@vms.huji.ac.il).

Pax6 is required for differentiation of glucagon-producing α -cells in mouse pancreas

Luc St-Onge, Beatriz Sosa-Pineda*, Kamal Chowdhury, Ahmed Mansouri & Peter Gruss

Max-Planck-Institut für Biophysikalische Chemie, 37077 Göttingen, Germany

The functional unit of the endocrine pancreas is the islet of Langerhans. Islets are nested within the exocrine tissue of the pancreas and are composed of α -, β -, δ - and γ -cells. β -Cells produce insulin and form the core of the islet, whereas α -, δ - and γ -cells are arranged at the periphery of the islet and secrete glucagon, somatostatin and a pancreatic polypeptide, respectively. Little is known about the molecular and genetic factors regulating the lineage of the different endocrine cells. Pancreas development is known to be abolished in *Pdx1*-mutant mice² and *Pax4* mutants lack insulin-producing β -cells³. Here we show that the paired-box gene *Pax6* is expressed during the early stages of pancreatic development and in mature endocrine cells. The pancreas of *Pax6* homozygous mutant mice lack glucagon-producing cells, suggesting that *Pax6* is essential for the differentiation of α -cells. As mice lacking *Pax4* and *Pax6* fail to develop any mature endocrine cells, we conclude that both *Pax* genes are required for endocrine fate in the pancreas.

Pax6, a member of the *Pax* gene family, is expressed in the developing eye, nose, pancreas and central nervous system⁴. Studies on the spontaneous mouse mutant *Small eye*, in which a point mutation in the *Pax6* locus results in a truncated protein, have shown that *Pax6* is essential for the formation of the lens placode⁵ and certain structures of the forebrain⁶. *Pax6* is also expressed in pancreatic endocrine cells⁷, but its function in these cells was unknown. We have now studied the role of *Pax6* during pancreas development in a new mouse mutant generated by replacing the start codon and the entire paired domain of *Pax6* with a β -galactosidase gene (Fig. 1a). Expression of β -galactosidase was therefore under the control of the endogenous *Pax6* promoter and could be used as a marker for cell lineage studies.

β -galactosidase activity in heterozygous embryos was identical to the *Pax6* expression pattern⁴. As we did not detect any ectopic expression in mutant mice, we considered that the β -galactosidase activity reflected *Pax6* expression. On day 9.0 of embryonic development (e9.0), strong β -galactosidase staining was observed in the telencephalon, diencephalon, neural tube, optic vesicle and pancreas (Fig. 1b). At this and later stages, *Pax6* expression was also detected in discrete cells of the pancreatic epithelium (Fig. 2c). *Pax6* expression was maintained in a subset of cells throughout development of dorsal and ventral pancreas (not shown) and became restricted to the islets of Langerhans in newborn animals (Fig. 2a–d). Hormonal production in embryonic and newborn pancreas was verified by indirect immunohistochemistry using monoclonal antibodies specific for insulin, glucagon, somatostatin and a pancreatic polypeptide (PP). Cells expressing both *Pax6* and glucagon could be detected as early as e10.5 (Fig. 1d), whereas at e15.5, *Pax6*-positive cells expressed either insulin or glucagon (not shown). *Pax6* expression was never detected in the exocrine tissue. In newborn animals, *Pax6*-positive cells synthesized insulin, glucagon, PP (Fig. 2a–c) or somatostatin (not shown), indicating that *Pax6* was expressed in mature α , β , δ and γ endocrine cells.

Homozygous *Pax6*–*LacZ* animals had a phenotype similar to

* Present address: St Jude Children's Research Hospital, Memphis, Tennessee 38101, USA.

homozygous *Small eye* mutants. They lacked eyes, olfactory bulb (not shown) and died a few minutes after birth. In the pancreas, β -galactosidase-positive cells were still detected in developing and newborn animals, but they failed to form distinct islets and remained disorganized throughout the exocrine tissue (compare Fig. 2a–d with e–h). Although positive cells produced insulin, PP (Fig. 2e,g,i,k) or somatostatin (not shown), endocrine cells normally found at the periphery of islets were mixed instead with β -cells (compare Fig. 2c with g,k). We detected no or very few glucagon-producing cells in mutant embryos or newborn animals (Fig. 2f,j). As glucagon production in the pancreas is attributed to α -cells, our results suggest an absence of mature α -cells in *Pax6* mutant animals. Exocrine tissue of mutant animals remained unaffected and still produced α -amylase (compare Fig. 2d with h). We could not determine whether the lack of α -cells corresponded to change in β -cell population because most endocrine cells in normal pancreas are β -cells and a conversion of α -cells into β -cells in mutant animals would not lead to any appreciable increase in their number.

Our results show that *Pax6* is expressed in mature endocrine cells in the islet of Langerhans of adult mouse pancreas. Expression of *Pax6* during the early stages of pancreatic development suggests that *Pax6* is also active in cells that will give rise to mature endocrine cells. Mutant animals lacking *Pax6* are deficient in glucagon-producing cells, indicating that *Pax6* is essential for α -cell differentiation. Furthermore, although the remaining β -, δ - and γ -cell populations in the pancreas of homozygous *Pax6*-mutant mice are still present, the islets of Langerhans are disorganized. In normal pancreas, the formation of islets is not clonal by origin⁸ but may be the result of cell migration and aggregation occurring late in development⁹. *Pax6* could be involved in organizing the endocrine cells into islets. Cell-adhesion molecules are thought to be important in islet formation^{10–12} and *Pax6* can bind to the promoter of at least one gene for a cell-adhesion molecule¹³. We are at present investigating whether a relation exists between *Pax6* and the expression of cell-adhesion molecules in the endocrine pancreas.

We have recently shown that the *Pax4* gene is required for β -cell differentiation because mutant mice do not develop insulin-

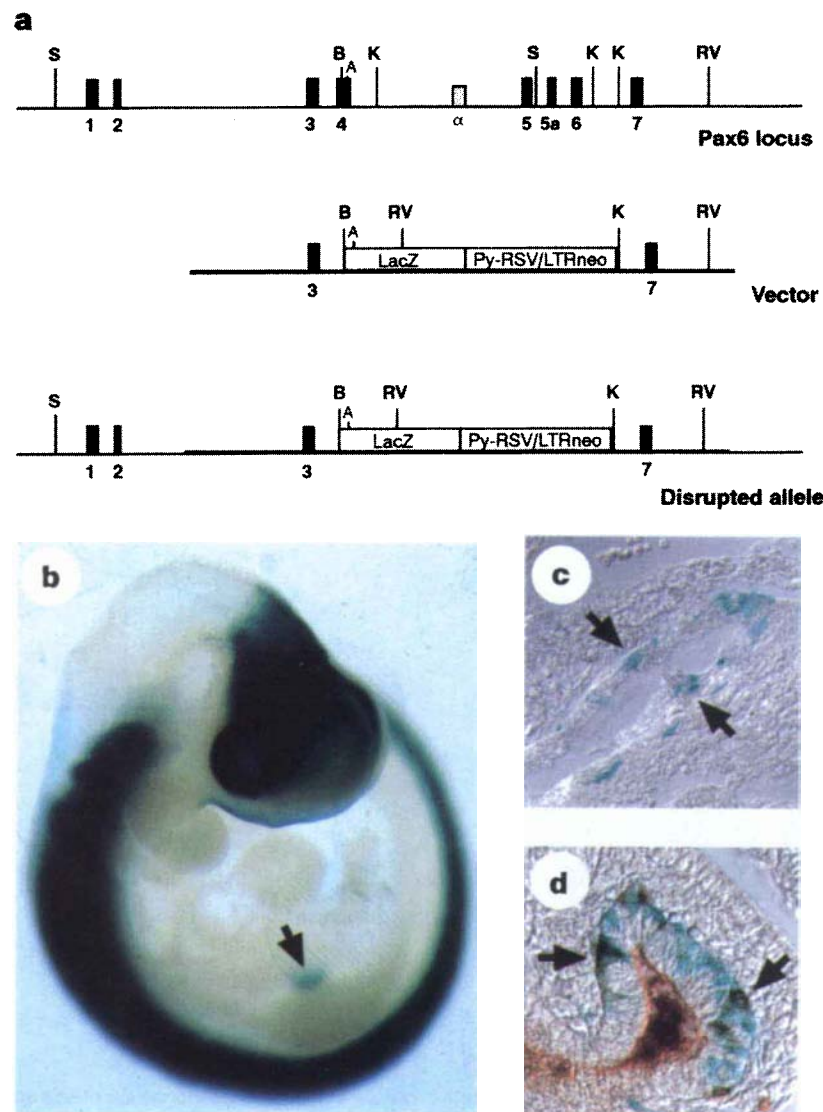


Figure 1 Targeted replacement of *Pax6* by the β -galactosidase gene and generation of heterozygous mice. **a**, Structure of wild-type and targeted *Pax6* loci. Deletion of the *Pax6* initiation codon and the entire paired domain was produced by inserting the β -galactosidase-neomycin cassette pGNA²⁰. Abbreviations: A, ATG; B, *Bam*HI; K, *Kpn*I; RV, *Eco*RV; S, *Spe*I. **b**, β -Galactosidase activity in e9.0 $+/-$ embryo. Strong staining is detected in the prosencephalon, diencephalon, optic

vesicle, neural tube and pancreas (arrowhead). **c**, Sagittal section of e9.0 (20 somites) $+/-$ embryo. Only a subset of cells in the pancreatic epithelium is positive for β -galactosidase expression. **d**, Sagittal section of the pancreatic evagination in e10.5 $+/-$ embryo. β -Galactosidase-positive cells coexpress glucagon (arrowheads). Magnification, $\times 400$.

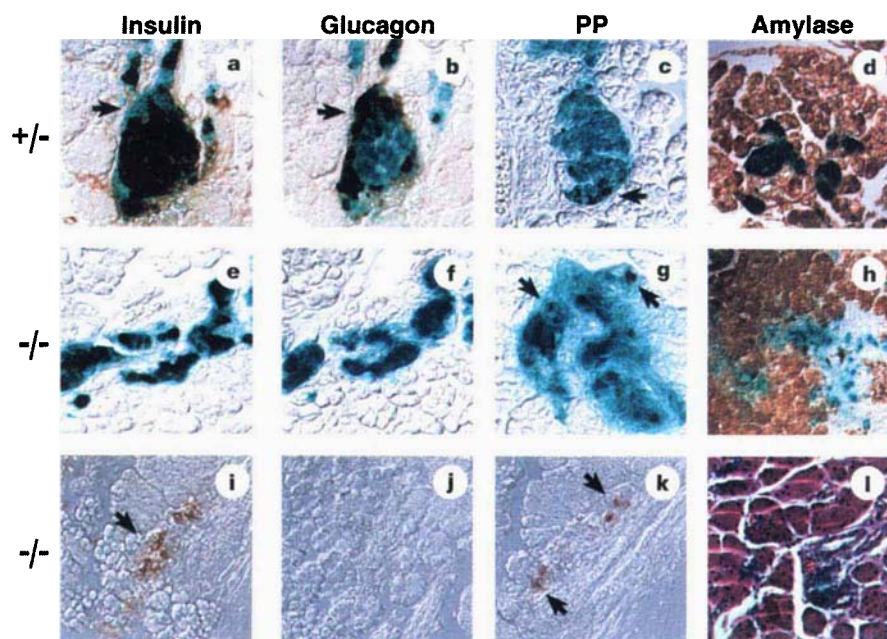


Figure 2 Analysis of β -galactosidase, insulin, glucagon, PP and amylase expression in newborn pancreas of $+/-$ and $-/-$ *Pax6*-deficient mice. β -Galactosidase staining is restricted to islet of Langerhans in $+/-$ mice (**a-d**). Blue-stained cells expressed either insulin (**a**), glucagon (**b**) or PP (**c**). Insulin-producing cells from the core of the islet; glucagon- and PP-producing cells are located at the periphery (arrowhead). In pancreata of $-/-$ mice stained (**e-g**)

and not stained (**i-k**) for β -galactosidase, cells expressed insulin (**e, i**) and PP (**g, k**) but not glucagon (**f, j**). Endocrine cells also failed to organize into islet-like structures. γ -Cells do not form a mantle around insulin cells (arrowhead). Exocrine tissue of $-/-$ animals possessed acinar cells (**l**) capable of producing α -amylase (**h**), like $+/-$ animals (**d**). **a, b, e, f**, and **i-k**, adjacent sections. Magnification is $\times 400$, except in **d, h**, where it is $\times 200$.

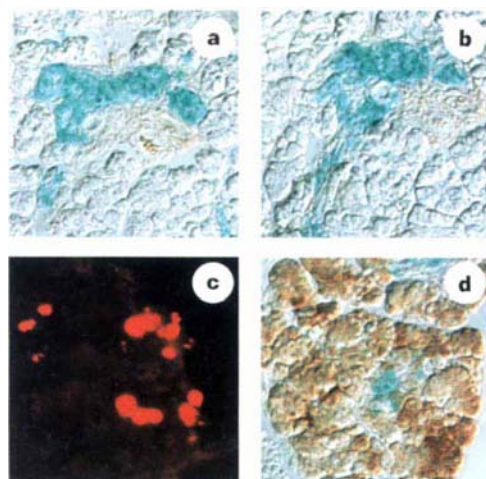


Figure 3 Analysis of β -galactosidase, insulin, glucagon and Pdx1 expression in newborn pancreas of mice homozygous for both *Pax6* and *Pax4*. Few β -galactosidase-positive cells were seen in mice lacking both *Pax4* and *Pax6* (**a-c**). These cells expressed neither insulin (**a**) nor glucagon (**b**), but a few Pdx1-positive cells could be detected in the pancreas (**c**). Exocrine tissue produced α -amylase (**d**), as did wild-type animals. Magnification in **a, b, d**, $\times 200$; in **c**, $\times 400$.

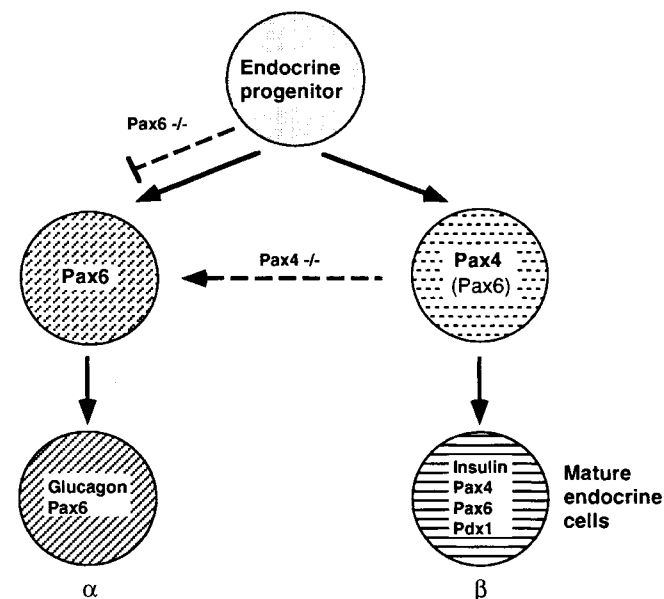


Figure 4 Proposed model for the role of *Pax4* and *Pax6* during differentiation of α - and β -cells during pancreatic development. Once the region of the foregut has acquired the ability to develop into a pancreas, *Pax* gene expression in the pancreatic epithelium at e9.0 defines cells into two populations: cells expressing both *Pax4* and *Pax6* will develop into mature insulin-producing β -cells; cells expressing only *Pax6* differentiate into glucagon-producing α -cells. Deletion of *Pax4* diverts the β -cell lineage into the α -cell lineage. Absence of *Pax6* eliminates the lineage for α -cells.

secreting cells³. We studied the fate of endocrine cells in the pancreas of newborn animals lacking both Pax6 and Pax4 by generating double mutants. In the pancreas of double homozygote newborn animals, a small population of β -galactosidase-positive cells was observed within the exocrine tissue that remained unaffected and produced α -amylase (Fig. 3d). Positive cells did not produce any endocrine hormones (Fig. 3a,b), whereas expression of the homeobox gene *Pdx1* could be detected in a few cells (Fig. 3c). β -Galactosidase-positive cells in double mutants may represent undifferentiated endocrine cells. As endocrine progenitors express both insulin and glucagon^{14,15}, further studies on the status of the progenitor cells in early double mutant embryos are required to determine when differentiation is arrested.

On the basis of our observations, we propose that Pax4 and Pax6 determine the cell fate of endocrine tissue during development (Fig. 4). All four endocrine cell types are thought to originate from a common pluripotent precursor that is derived from the endoderm. Once the primitive gut epithelium has acquired the potential to form the pancreas, the onset of Pax gene expression defines the lineage of the different endocrine cells. Cells expressing both Pax4 and Pax6 differentiate into mature β -, δ -, γ -cells and cells expressing only Pax6 differentiate into α -cells. Absence of Pax4 diverts all cells into the α -cell lineage as Pax6 remains present. This is supported by the observation that homozygous Pax4-mutant mice lack β -cells but have a larger than normal α -cell population³. Deletion of Pax6 eliminates the α -cell lineage altogether or diverts cells to the β -cell lineage. The absence of both Pax4 and Pax6 results in cells that fail to differentiate into any mature endocrine cells. Thus, both Pax genes are required for endocrine cell differentiation during pancreatic development. □

Methods

Animals. Pax6-LacZ mice were generated by homologous recombination in the R1 ES cell line according to standard procedures^{16,17} and positive mutant clones were used to produce chimaeric animals by the aggregation technique¹⁸. Chimaeras were then mated to NMRI mice for germline transmission.

β -Galactosidase staining. β -Galactosidase activity was verified after fixing embryos and pancreas for 30 min in 2% formaldehyde/0.02% glutaraldehyde in PBS and staining overnight at 30 °C in a PBS solution containing 0.1% X-gal, 2 mM MgCl₂, 5 mM K₃Fe(CN)₆, 5 mM K₄Fe(CN)₆, 0.2% Nonidet P-40 and 2.5 mM deoxycholic acid. After staining, tissues were embedded in paraffin and 10- μ m sections were prepared.

Immunohistochemistry. Primary antibodies mouse anti-insulin (Sigma), mouse anti-glucagon (Sigma), rabbit anti-PP (Dako), rabbit anti-somatostatin (Dako) and rabbit anti-human α -amylase (Sigma) were applied on paraffin sections after β -galactosidase staining and detected with a secondary horseradish peroxidase antibody as described³. Rabbit anti-PDX1 (ref. 19) was applied to cryostat sections and revealed with a secondary anti-rabbit TRITC antibody (Sigma).

Received 13 January; accepted 21 March 1997.

- Slack, J. M. Developmental biology of the pancreas. *Development* **121**, 1569–1580 (1995).
- Jonsson, J., Carlsson, L., Edlund, T. & Edlund, H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606–609 (1994).
- Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G. & Gruss, P. The Pax4 gene is essential for differentiation of insulin-producing β -cells in mammalian pancreas. *Nature* **386**, 399–402 (1997).
- Walther, C. & Gruss, P. Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development* **113**, 1435–1449 (1991).
- Grindley, J. C., Davidson, D. R. & Hill, R. E. The role of Pax6 in eye and nasal development. *Development* **121**, 1433–1442 (1995).
- Stoykova, A., Fritsch, R., Walter, C. & Gruss, P. Forebrain patterning defect in *Small eye* mutant mice. *Development* **122**, 3453–3465 (1996).
- Turque, N., Plaza, S., Radvanyi, F., Carriere, C. & Saule, S. Pax-QNR/PAX-6, a paired box- and homeobox-containing gene expressed in neurons, is also expressed in pancreatic endocrine cells. *Mol. Endocrinol.* **8**, 929–938 (1994).
- Deltour, L. et al. Polyclonal origin of pancreatic islets in aggregation mouse chimaera. *Development* **112**, 1115–1121 (1991).
- Herrera, P. L. et al. Embryogenesis of the murine endocrine pancreas; early expression of pancreatic polypeptide gene. *Development* **113**, 1257–1265 (1991).
- Dahl, U., Sjödin, A. & Semb, H. Cadherins regulate aggregation of pancreatic β -cells *in vivo*. *Development* **122**, 2895–2902 (1996).
- Möller, C. J. et al. Differential expression of neural cell adhesion molecule and cadherins in pancreatic islets, glucagonomas, and insulinomas. *Mol. Endocrinol.* **6**, 1332–1342 (1992).
- Rouiller, D. G., Cirrulli, V. & Halban, P. Uvomorulin mediates calcium-dependent aggregation of islet

- cells, whereas calcium-independent cell adhesion molecules distinguish between islet cells types. *Dev. Biol.* **148**, 233–242 (1991).
- Chalepakos, G., Wijnholds, J., Giese, P., Schachner, M. & Gruss, P. Characterization of Pax-6 and Hoxa-1 binding to the promoter region of the neural cell adhesion molecule L1. *DNA Cell Biol.* **13**, 891–900 (1994).
- Gittes, G. K. & Rutter, W. J. Onset of cell-specific gene expression in the developing mouse pancreas. *Proc. Natl Acad. Sci. USA* **89**, 1128–1132 (1992).
- Teitelman, G., Alpert, S., Polak, J. M., Martinez, A. & Hanahan, D. Precursor cells of mouse endocrine pancreas coexpress insulin, glucagon and the neuronal protein tyrosine hydroxylase and neuropeptide Y, but not pancreatic polypeptide. *Development* **118**, 1031–1039 (1993).
- Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
- Wurst, W. & Joyner, A. L. in *Gene Targeting: A Practical Approach* (ed. Joyner, A. L.) 33–61 (Oxford University Press, Oxford, 1993).
- Nagy, A. & Rossant, J. in *Gene Targeting: A Practical Approach* (ed. Joyner, A. L.) 145–179 (Oxford University Press, Oxford, 1993).
- Ohlsson, H., Karlsson, K. & Edlund, T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J.* **12**, 4251–4259 (1993).
- Le Mouellac, H., Lallemand, Y. & Brûlet, P. Targeted replacement of the homeobox gene Hox-3.1 by the *Escherichia coli* LacZ in mouse chimeric embryos. *Proc. Natl Acad. Sci. USA* **87**, 4712–4716 (1990).

Acknowledgements. We thank H. Edlund for the Pdx1 antibody; M. Kessel for critical reading of the manuscript; H. Brinkmann, K. Rollert and J. Krull for technical assistance; and R. Altschäffler for photographic artwork. L.S.O. was supported by a fellowship from the National Cancer Institute of Canada. This work was supported by the Max-Planck Society.

Correspondence and requests for material should be addressed to P.G. (e-mail: PGruss@gwdg.de).

Expression cloning and characterization of a renal electrogenic Na⁺/HCO₃⁻ cotransporter

Michael F. Romero^{*}, Matthias A. Hediger[†],
Emile L. Boulpaep^{*} & Walter F. Boron^{*}

^{*} Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

[†] Renal Division, Brigham and Women's Hospital, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Bicarbonate transporters are the principal regulators of pH in animal cells, and play a vital role in acid–base movement in the stomach, pancreas, intestine, kidney, reproductive system and central nervous system. The functional family of HCO₃⁻ transporters includes Cl⁻–HCO₃⁻ exchangers, three Na⁺/HCO₃⁻ cotransporters^{1–3}, a K⁺/HCO₃⁻ cotransporter^{4,5}, and a Na⁺-driven Cl⁻–HCO₃⁻ exchanger^{6,7}. Molecular information is sparse on HCO₃⁻ transporters, apart from Cl⁻–HCO₃⁻ exchangers ('anion exchangers'), whose complementary DNAs were cloned several years ago^{8–11}. Attempts to clone other HCO₃⁻ transporters, based on binding of inhibitors, protein purification or homology with anion exchangers, have so far been unsuccessful. Here we monitor the intracellular pH and membrane voltage in *Xenopus* oocytes to follow the expression of the most electrogenic transporter known: the renal 1:3 electrogenic Na⁺/HCO₃⁻ cotransporter from the salamander *Ambystoma tigrinum*. We now report the successful cloning and characterization of a cDNA encoding a cation-coupled HCO₃⁻ transporter. The encoded protein is 1,035 amino acids long with several potential membrane-spanning domains. We show that when it is expressed in *Xenopus* oocytes, this protein is electrogenic, Na⁺ and HCO₃⁻ dependent, and blocked by the anion-transport inhibitor DIDS, and conclude that it is the renal electrogenic sodium bicarbonate cotransporter (NBC).

Our expression-cloning strategy used a new assay and source of poly(A)⁺ RNA. For the assay, we simultaneously monitored membrane voltage (V_m) and intracellular pH (pH_i) with voltage- and pH-sensitive microelectrodes in *Xenopus laevis* oocytes. The protocol was to equilibrate the oocyte with CO₂/HCO₃⁻ and then remove extracellular Na⁺. Because the cotransporter exports more HCO₃⁻

than Na^+ , this causes a positive shift in V_m and a fall in pH_i . Injecting oocytes with poly(A)⁺ RNA from rabbit kidney cortex failed to elicit the expected V_m and pH_i signals. Reasoning that an amphibian oocyte might recognize and translate amphibian RNA better, we turned to the salamander, in which the cotransporter was first described¹. We then obtained the expected V_m and pH_i signals in oocytes injected with poly(A)⁺ RNA obtained from *Ambystoma* kidney.

When water-injected, control *Xenopus* oocytes are exposed to $\text{CO}_2/\text{HCO}_3^-$, pH_i falls and the cells depolarize (Fig. 1a). Subsequently removing external Na^+ leads to a small hyperpolarization and no pH_i change. In poly(A)⁺-RNA-injected oocytes, however, removing Na^+ produces a transient depolarization and acidification (Fig. 1b), a response characteristic of electrogenic $\text{Na}^+/\text{HCO}_3^-$ cotransporters¹. Although the depolarization is strong, the pH_i decrease is less marked, but consistently observed (Fig. 1c). Poly(A)⁺-RNA-injected oocytes treated with 400 μM DIDS behave as water-injected controls (not shown).

Screening an *Ambystoma* kidney cDNA library for electrogenic $\text{Na}^+/\text{HCO}_3^-$ cotransporter activity yielded a 4.1-kilobase (kb) cDNA encoding a new 1,025-amino-acid protein (which we name NBC, for $\text{Na}^+/\text{bicarbonate}$ cotransporter) (Fig. 2a). To determine whether our clone had the true start methionine, we used poly-

merase chain reaction (PCR) to amplify NBC-specific 5' fragments from the library master plates. Sequencing clones from three master plates, we identified an additional 186 base pairs (bp) on the 5' end, yielding ten additional amino acids. Thus, NBC has a total of 1,035 amino acids and a 5' untranslated region of 156 bp. Comparing the entire deduced amino-acid sequence of NBC with proteins in GenBank reveals no major homology to other known proteins, indicating that NBC is the first member of a new family of Na^+ -coupled HCO_3^- transporters. However, there is some homology to the anion exchangers (~25–30% amino-acid identity over the entire sequence, allowing for gaps). At a site homologous to the anion exchanger DIDS-binding motif, KL(X)K, the NBC sequence is KMIK. NBC has a KLKK sequence where the anion exchangers have (K/R)(L/M)(X)K (Fig. 2a). NBC's hydropathy plot (Fig. 2b) is remarkably similar to that of the anion exchangers. A dendrogram (Fig. 2c) shows that NBC and the anion exchangers define a new superfamily of HCO_3^- transporters. We suggest that NBC has ten putative membrane-spanning domains, with a large 5–6 exofacial loop with several consensus N-glycosylation sites (Fig. 2d).

Using high-stringency northern blot analysis, we examined the localization of NBC mRNA in 19 *Ambystoma* tissues. A ~4.2-kb transcript is found in the kidney, as well as bladder, brain, small intestine, large intestine, and eye (Fig. 3), consistent with the known

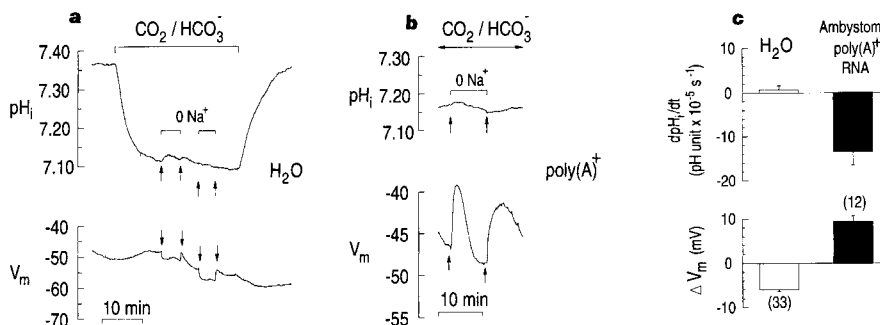


Figure 1 Expression-cloning assay in *Xenopus* oocytes. **a**, Control oocyte injected with 50 nl water. The external solution (pH 7.5) was switched from ND96 ($\text{CO}_2/\text{HCO}_3^-$ free) to a solution buffered with 1.5% $\text{CO}_2/10 \text{ mM HCO}_3^-$. Na^+ was removed in the presence of the $\text{CO}_2/\text{HCO}_3^-$. Similar results were obtained upon removing Na^+ from ND96 (not shown). **b**, Oocyte injected with *Ambystoma* kidney poly(A)⁺ RNA (25 ng in 50 nl). $\text{CO}_2/\text{HCO}_3^-$ was added before and maintained

during the portion of the experiment shown (→). The protocol was the same as in **a**. Expression was first evident at day 6 after injection and continued until day 13. **c**, Summary of initial rates of pH_i change ($d\text{pH}_i/dt$) and ΔV_m resulting from Na^+ removal in 1.5% $\text{CO}_2/10 \text{ mM HCO}_3^-$ for oocytes injected with water or *Ambystoma* kidney poly(A)⁺ RNA.

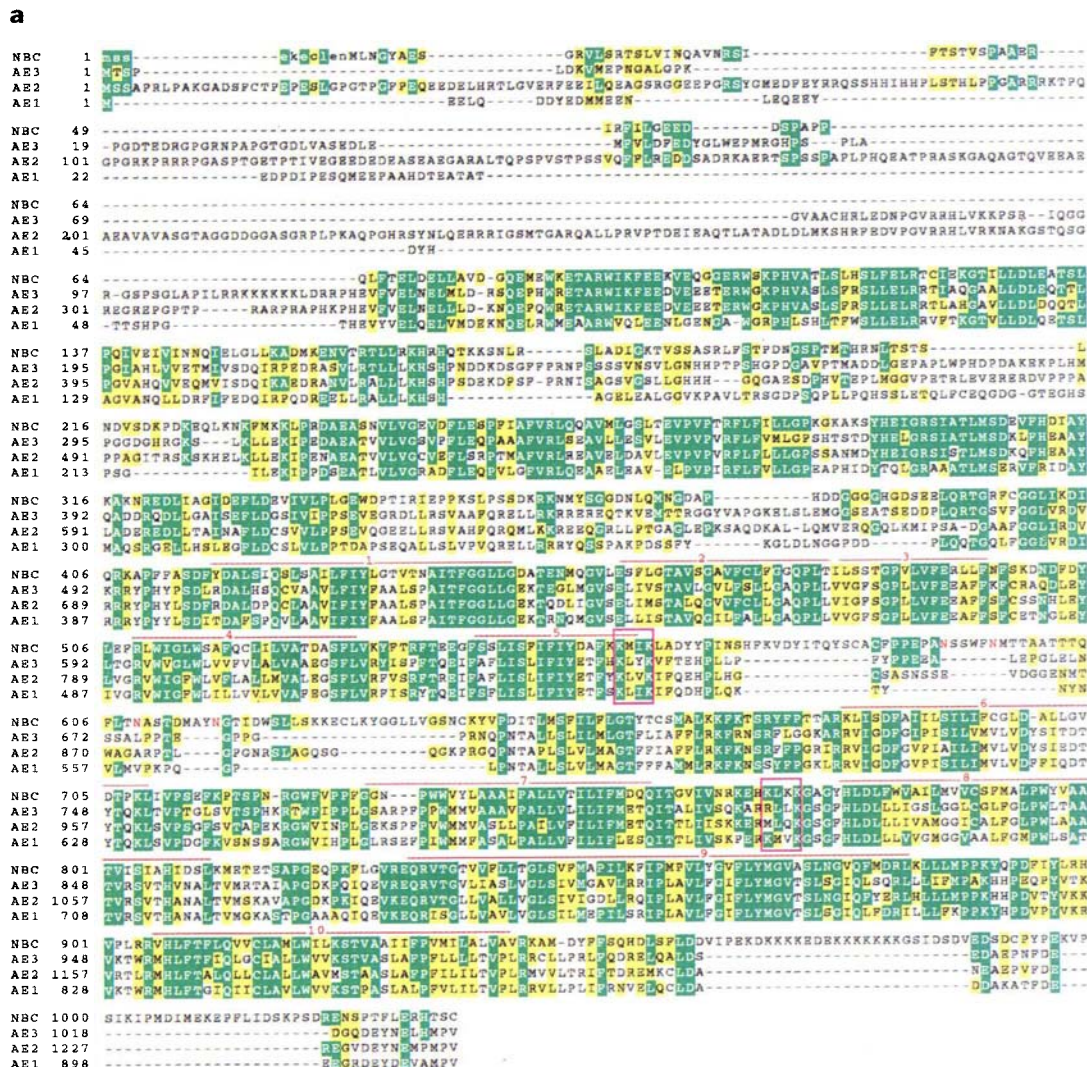
Figure 2 NBC sequence. **a**, Multiple sequence alignment of NBC with the anion exchangers (AEs). A 4,078-bp cDNA encoding the renal electrogenic Na^+ bicarbonate cotransporter (NBC) was isolated by screening a size-selected (3.5–5.0 kb) *Ambystoma*-kidney cDNA library for electrogenic cotransport of Na^+ and HCO_3^- (Fig. 1). The cDNA has a poly(A)⁺ tail, an open reading frame from nucleotides 33 to 3107 encoding a 1,025-amino-acid protein (capital), and a large 3'-untranslated region (~1 kb; not shown). Ten additional amino acids 5' to the original start methionine (M) are indicated in lower case (1,035 amino acids total). Membrane-spanning domains (MSDs) of NBC are indicated by a red line over the NBC sequence. In the multiple sequence alignment, AE sequences (GenBank accession numbers: S03074, S21086, A42497) identical to NBC are highlighted in green with reverse type, and areas of homology (acidic, basic, hydrophobic or polar) to NBC are highlighted in yellow. In the region homologous to the AE consensus DIDS-binding motif KL(X)K (where X is I, V or Y)¹⁰, NBC has the sequence KMIK (shaded area, residues 558–561), suggesting a modified motif of K(Y)(X)K, where Y is M or L. NBC contains the AE consensus DIDS-binding motif, KLKK, at residues 768–771 (boxed sequence). Putative N-linked glycosylation sites in NBC are in red. NBC is homologous to several EST (expressed sequence tag) sequences (for example, W31917, W39298, N27899 and N58147) and a genomic *Caenorhabditis elegans* clone (F52B5.1). The GenBank accession number of NBC is AF001958. **b**, Hydropathy plot of NBC. Ten MSDs (that is, 10

hydrophobic regions) are predicted from the primary amino-acid sequence using a Kyte-Doolittle algorithm (window size, 18 amino acids). A large exofacial loop (5.6-EFL) is predicted between MSDs 5 and 6. For the AEs, putative MSDs 9 and 10 are extensive enough to span the membrane twice²⁶, consistent with 12 MSDs; others have suggested as many as 14 (refs 26, 27). As for the AEs, both the N and C termini are constrained to be intracellular. **c**, Dendrogram showing the percentage divergence of the amino-acid sequences of NBC (aNBC for *Ambystoma* NBC) and the most homologous AEs (GenBank accession numbers: S03074, S21086, A42497). The divergence is indicated by the total length of the horizontal line segments from one label (for example, 'aNBC') to another. NBC is 35, 33 and 34% identical to AE3, AE2 and AE3, respectively. **d**, Membrane model of NBC protein. Putative MSDs are indicated by numbered rectangles. DIDS-binding motifs are indicated as green diamonds. Of 8 consensus N-linked glycosylation sites, only 4 (N-glyc, in red) are predicted to the extracellular (amino acids 591, 596, 609 and 617; not 34, 160, 209 and 496). Ser982, predicted to be intracellular, is the only consensus protein kinase A (PKA; pink triangle) site. Of the 15 consensus sites for protein kinase C (PKC; blue circles), 12 are predicted to be intracellular (Ser 18, Thr 85, Thr 173, Thr 206, Ser 219, Thr 345, Ser 357, Thr 394, Thr 706, Ser 810, Ser 1000, Ser 1020), and 3 to be extracellular (Ser 626, Thr 671, Thr 678). Charged residues closely associated with MSDs, or those at a high density, are indicated by plus or minus signs.

roles of $\text{Na}^+/\text{HCO}_3^-$ cotransport in these tissues. A smaller transcript of ~ 1.8 kb appears in kidney, bladder, skeletal muscle and heart.

Expression of the pure NBC clone is shown in Fig. 4. A feature of cells injected with NBC rather than poly(A)⁺ RNA is the immediate hyperpolarization (53.5 ± 2.3 mV; $n = 27$) caused by applying $\text{CO}_2/\text{HCO}_3^-$ (Fig. 4a) due to the inward (reverse) flux of Na^+ and HCO_3^- . Removing Na^+ forces NBC to operate in the outward direction (normal for the renal proximal tubule), depolarizing and acidifying the cell. 200 μM DIDS blocks these V_m and pH_i changes. Thus, NBC is Na^+ dependent and blocked by DIDS.

Next, we tested NBC's HCO_3^- dependence. Because the cotransporter is pH_i -sensitive¹², we matched the acidification produced by $\text{CO}_2/\text{HCO}_3^-$ with one caused by butyrate. As shown in Fig. 4b, removing Na^+ produces the characteristic depolarization (51.0 ± 2.4 mV; $n = 34$) when the oocyte is acidified with CO_2 , but a hyperpolarization when the oocyte is acidified with butyrate. As is the case with lowering the Na^+ concentration outside the cell, lowering extracellular HCO_3^- concentration at a fixed CO_2 concentration causes a large depolarization (51.6 ± 4.0 mV; $n = 7$) because of $\text{Na}^+/\text{HCO}_3^-$ efflux (Fig. 4c). Conversely, raising extracellular



HCO_3^- concentration hyperpolarizes the cell ($37.0 \pm 1.7 \text{ mV}$; $n = 9$).

Finally, we evaluated the electrogenic nature of NBC in voltage-clamp experiments. Adding $\text{CO}_2/\text{HCO}_3^-$ elicits an outward current (Fig. 4d), consistent with $\text{Na}^+/\text{HCO}_3^-$ influx. Removing Na^+ elicits inward currents of similar magnitude, consistent with $\text{Na}^+/\text{HCO}_3^-$ efflux.

Our NBC clone encodes the renal electrogenic $\text{Na}^+/\text{HCO}_3^-$ cotransporter that mediates HCO_3^- efflux and presumably has a $\text{Na}^+:\text{HCO}_3^-$ stoichiometry of 1:3 (refs 13, 14). A functionally related

cotransporter, which mediates both HCO_3^- influx and efflux and has a stoichiometry of 1:3, has also been identified in retinal Müller cells^{15,16}. Other cotransporters, presumably related to NBC at the molecular level, mediate HCO_3^- influx. Cotransporters with $\text{Na}^+:\text{HCO}_3^-$ stoichiometries of 1:2 are found in glia², liver¹⁷⁻¹⁹, pancreas²⁰ and colon²¹. A cotransporter with a stoichiometry of 1:1 is present in heart³. Consequently, if the cotransporter encoded by our NBC clone proves to have a 1:3 stoichiometry, and if clones encoding $\text{Na}^+/\text{HCO}_3^-$ cotransporters of different stoichiometries are identified, we propose the name NBC-3 for the present clone.

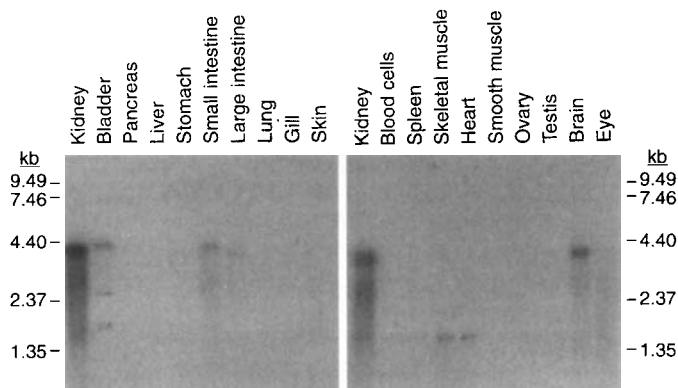


Figure 3 High-stringency northern blot analysis of poly(A)⁺ RNA from *Ambystoma* tissues probed with ³²P-labelled NBC-cDNA and exposed for 21 h. After 10 h exposure (not shown), a 4.2-kb transcript was found only in the kidney lanes. After

21 h exposure, similar transcripts are seen in bladder, small intestine, large intestine, brain and eye.

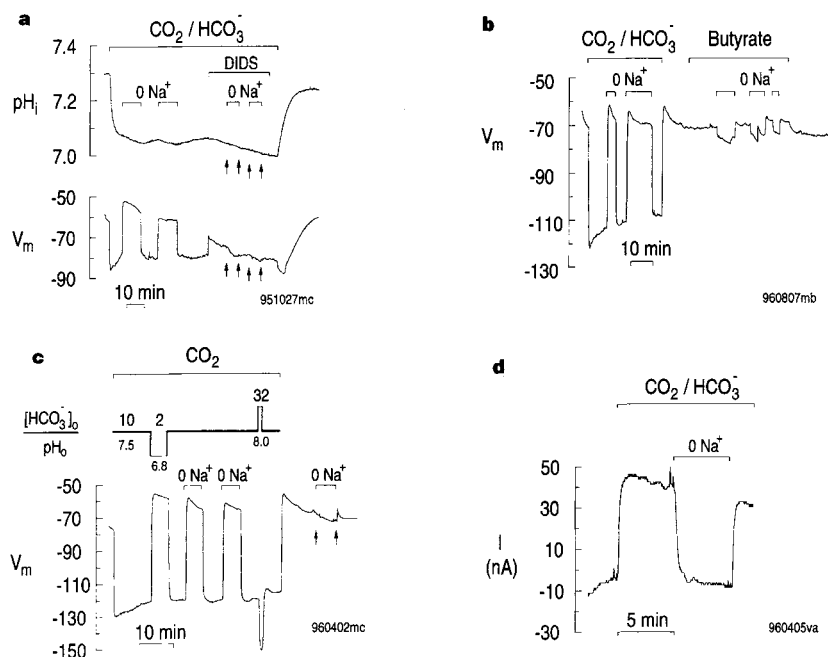


Figure 4 Physiology of NBC clone expressed in oocytes. In contrast to the expression of poly(A)⁺ RNA (Fig. 1), expression of NBC cRNA (10 ng in 50 nl) was first obvious on day 3 after injection and continued until at least day 13. **a**, DIDS (4,4'-diisothiocyanato-2,2'-stilbene disulphonate) sensitivity. The external solution was switched from ND96 solution ($\text{CO}_2/\text{HCO}_3^-$ free) to 1.5% $\text{CO}_2/10 \text{ mM HCO}_3^-$ (pH 7.5). Na^+ was then removed four times in $\text{CO}_2/\text{HCO}_3^-$. The last two Na^+ abstractions were made in the presence of $200 \mu\text{M}$ DIDS. The response to Na^+ removal in the presence of DIDS was similar to that in water-injected oocytes in the absence of DIDS (Fig. 1a). **b**, HCO_3^- dependence. From a resting value of 7.2–7.4 in ND96, pH_i decreased to ~ 7.0 within 10 min of adding either 1.5% $\text{CO}_2/10 \text{ mM HCO}_3^-$ or 10 mM butyrate (not shown). Na^+ was removed five times, twice in $\text{CO}_2/\text{HCO}_3^-$, and three times in butyrate. The hyperpolarizing response to Na^+ removal

in butyrate is similar to that observed in water-injected oocytes in the presence of $\text{CO}_2/\text{HCO}_3^-$ (Fig. 1a). **c**, Changing extracellular $[\text{HCO}_3^-]_o$ at a fixed CO_2 concentration by altering extracellular pH (pH_o). Maintaining a P_{CO_2} of 1.5%, we changed $[\text{HCO}_3^-]_o$ from 10 to 2 mM (pH_o was 7.5–6.8). After two zero- Na^+ pulses, we changed $[\text{HCO}_3^-]_o$ from 10 to 32 mM (pH_o , 7.5–8.0). **d**, NBC currents. The oocyte was voltage-clamped at -60 mV as we switched from ND96 to 1.5% $\text{CO}_2/10 \text{ mM HCO}_3^-$. The $\sim 50 \text{ nA}$ outward current elicited by $\text{CO}_2/\text{HCO}_3^-$ corresponds to a $\sim 50 \text{ mV}$ hyperpolarization observed in unclamped cells; the $\sim 50 \text{ nA}$ inward current elicited by Na^+ removal corresponds to a $\sim 50 \text{ mV}$ depolarization observed in unclamped cells. Thus, NBC expression does not significantly change the native oocyte resistance of $\sim 1 \text{ M}\Omega$.

In summary, the newly cloned NBC, when expressed in *Xenopus* oocytes, has the physiological properties of the renal electrogenic $\text{Na}^+/\text{HCO}_3^-$ cotransporter: electrogenicity, Na^+ dependence, HCO_3^- dependence, and sensitivity to DIDS. Now that this Na^+ -coupled HCO_3^- transporter has been cloned, it may assist in the molecular identification of other cation-coupled HCO_3^- transporters. Cloning of other $\text{Na}^+/\text{HCO}_3^-$ cotransporters should provide insight into the crucial roles they play in the heart, stomach, pancreas and intestine. In the case of the kidney, knowing the NBC cDNA sequence should help us to understand disorders of excessive Na^+ retention such as hypertension, or impaired HCO_3^- reclamation, as in renal tubular acidosis. In the brain, identifying NBC-like cDNAs may help to clarify how electrogenic $\text{Na}^+/\text{HCO}_3^-$ cotransport modulates neuronal activity. Comparison of the cDNA sequences of the NBC cotransporter and the anion exchangers may help elucidate the mechanism of transporter function. □

Methods

RNA isolation and cDNA library. Poly(A)⁺ RNA was isolated from whole kidneys of 50 *Ambystoma*, using guanidinium isothiocyanate and phenol/chloroform extraction²⁷, and an aliquot was injected into *Xenopus* oocytes for expression²³. The remaining poly(A)⁺ RNA was size-fractionated by preparative gel electrophoresis^{23–25} to yield a sized RNA sample with an enriched response. DNAs were synthesized from this size-selected poly(A)⁺ RNA (3.5–5 kb) using a SuperScript plasmid system (GibcoBRL). cDNAs were size-selected, then directionally ligated into pSport 1, and the resulting plasmids used to transform *Escherichia coli*. About 8,000 clones were plated onto 15 master plates, plasmids were isolated, then cRNAs were transcribed *in vitro* and injected into *Xenopus* oocytes for expression. Positive plates (3 of 15) were identified by a depolarization and pH_i decrease elicited by Na^+ removal in $\text{CO}_2/\text{HCO}_3^-$ (Fig. 1b). From one positive plate, 538 colonies were picked and plated onto 16 submaster plates in a grid pattern. A positive submaster plate was identified as described. Successively smaller pools of cDNA clones were screened until a single clone, named NBC, was identified (Fig. 4). Sequencing was performed by the Keck Biotechnology Resource Laboratory (Boyer Center for Molecular Medicine, Yale University).

PCR protocol. We used PCR to identify 5' regions of clones from the 15 *Ambystoma* library master plates. Using T7 (primer to pSport 1) and an antisense sequencing primer (specific for NBC), *Taq* polymerase, and dNTPs, we did a hot-start PCR with 30 cycles at 94°C (30 s), 55°C (45 s) and 72°C (45 s). PCR products (13 of 15 master plates) were subcloned using a TA cloning kit (Invitrogen) and sequenced.

Northern analysis. Poly(A)⁺ RNA was isolated with Trizol reagent (GibcoBRL) from tissues obtained from 25 *Ambystoma tigrinum*, followed by oligo(dT) selection with oligo(dT) cellulose (GibcoBRL) or FastTracks (Ambion). Tissue poly(A)⁺ RNA (2 µg) was electrophoresed in a denaturing (50% formaldehyde) MOPS buffer, and then blotted and ultraviolet-crosslinked to a Zeta-Probe nylon membrane (BioRad). The blot was probed with ³²P-random-hexamer-primed DNA (GibcoBRL) from the open reading frame of NBC using ExpressHyb (Clontech).

Sequence analysis. Multiple sequence alignments were performed using the Clustal method and the PAM250 residue-weight table (DNASTAR program, Lasergene, Madison, WI). Residue similarities were determined according to the following scheme: acid, DE; basic, HKR; hydrophilic, AFILMPVW; and hydrophobic, CGNQSTY. The dendrogram '% divergence' was calculated as: divergence (*i, j*) = [100 × distance (*i, j*)]/(total distance), where distance (*i, j*) = sum (residual distances) + (gaps × gap penalty) + (gap residues × gap-length penalty); similarity (*i, j*) = [100 × sum of matches]/[length – gap residue(*i*) – gap residue(*j*)], where *i, j* are the compared sequences. The dendrogram and '% similarity' calculations ignore poor matches on the ends of either sequence, so the values are misleadingly high. For example, consider the sequences CDEF, CDE, DEF and CEF: CDEF is calculated to be 100% similar to both CDE and DEF, but only 67% similar to CEF. By inspection, it is clear that the two examples of 100% similarity are overestimates, because 25% of the respective sequences are ignored for the comparison.

Oocyte experimental solutions. The $\text{CO}_2/\text{HCO}_3^-$ -free solution ND96 contained 96 mM NaCl, 2 mM KCl, 1 mM MgCl_2 , 1.8 mM CaCl_2 and 5 mM HEPES

(pH 7.5). In $\text{CO}_2/\text{HCO}_3^-$ -equilibrated solutions, 10 mM NaHCO_3 replaced 10 mM NaCl. Solutions contained 100 µM ouabain to block the oocyte's native Na^+/K^+ pump. In 0- Na^+ solutions, choline replaced Na^+ . In butyrate solutions, 10 mM sodium butyrate replaced 10 mM NaCl.

Oocyte electrophysiology. 50 nl water (control) or RNA solution (25 ng *Ambystoma* poly(A)⁺ RNA) was injected into *Xenopus* oocytes. Voltage electrodes, made from fibre-capillary borosilicate and filled with 3 M KCl, had resistances of 1–10 MΩ. pH electrodes were pulled similarly and silanized with bis-(dimethylamino)-dimethylsilane (Fluka). Their tips were filled with hydrogen-ionophore-1 cocktail B (Fluka), and back-filled with phosphate buffer (pH 7.0). Electrodes were connected to a high-impedance electrometer (WPI-FD223 for pH_i and *V*_m experiments) and output data acquired by computer. pH electrodes had slopes of –54 to –57 mV per pH unit between pH 6.0 and 8.0. For voltage-clamp experiments (Warner Instrument Co., oocyte clamp), electrodes were filled with 3 M KCl and had resistances of 0.5–1.0 MΩ. Oocytes were clamped at –60 mV; the resulting currents were sampled at 0.5 Hz and filtered at 25 Hz (8-pole Bessel filter; Frequency Devices).

Received 19 September 1996; accepted 14 March 1997.

- Boron, W. F. & Boulpaep, E. L. Intracellular pH regulation in the renal proximal tubule of the salamander: basolateral HCO_3^- transport. *J. Gen. Physiol.* **81**, 53–94 (1983).
- Deitmer, J. W. & Schlue, W.-K. The regulation of intracellular pH by identified glial cells and neurones in the central nervous system of the leech. *J. Physiol. (Lond.)* **388**, 261–283 (1987).
- Dart, C. & Vaughan-Jones, R. D. $\text{Na}^+/\text{HCO}_3^-$ symport in the sheep cardiac Purkinje fibre. *J. Physiol. (Lond.)* **451**, 365–385 (1992).
- Hogan, E. M., Cohen, M. A. & Boron, W. F. K^+ - and HCO_3^- -dependent acid-base transport in squid giant axons: I. Base efflux. *J. Gen. Physiol.* **106**, 821–844 (1995).
- Hogan, E. M., Cohen, M. A. & Boron, W. F. K^+ - and HCO_3^- -dependent acid-base transport in squid giant axons: II. Base influx. *J. Gen. Physiol.* **106**, 845–862 (1995).
- Russell, J. M. & Boron, W. F. Role of chloride transport in regulation of intracellular pH. *Nature* **264**, 73–74 (1976).
- Thomas, R. C. The role of bicarbonate, chloride and sodium ions in the regulation of intracellular pH in snail neurones. *J. Physiol. (Lond.)* **273**, 317–338 (1977).
- Kopito, R. R. & Lodish, H. F. Primary structure and transmembrane orientation of the murine anion exchange protein. *Nature* **316**, 234–238 (1985).
- Alper, S. L., Kopito, R. R., Libresco, S. M. & Lodish, H. F. Cloning and characterization of a murine band 3-related cDNA from kidney and from a lymphoid cell line. *J. Biol. Chem.* **263**, 17092–17099 (1988).
- Kopito, R. R. *et al.* Regulation of intracellular pH by a neuronal homolog of the erythrocyte anion exchanger. *Cell* **59**, 927–937 (1989).
- Kudrycki, K. E., Newman, P. R. & Shull, G. E. cDNA cloning and tissue distribution of mRNAs for two proteins that are related to the band 3 $\text{Cl}^-/\text{HCO}_3^-$ exchanger. *J. Biol. Chem.* **265**, 462–471 (1990).
- Soleimani, M., Hattabaugh, Y. J. & Bizal, G. L. pH sensitivity of the $\text{Na}^+/\text{HCO}_3^-$ cotransporter in basolateral membrane vesicles isolated from rabbit kidney cortex. *J. Biol. Chem.* **267**, 18349–18355 (1992).
- Yoshitomi, K., Burckhardt, B.-C. & Frömter, E. Rheogenic sodium-bicarbonate cotransport in the peritubular cell membrane of rat renal proximal tubule. *Pflügers Arch.* **405**, 360–366 (1985).
- Soleimani, M., Grassl, S. M. & Aronson, P. S. Stoichiometry of $\text{Na}^+/\text{HCO}_3^-$ cotransport in basolateral membrane vesicles isolated from rabbit renal cortex. *J. Clin. Invest.* **79**, 1276–1280 (1987).
- Newman, E. A. & Astion, M. L. Localization and stoichiometry of electrogenic sodium bicarbonate cotransport in retinal glial cells. *Glia* **4**, 424–428 (1991).
- Newman, E. A. Acid efflux from retinal glial cells generated by sodium bicarbonate cotransport. *J. Neurosci.* **16**, 159–168 (1996).
- Fitz, J. G., Persico, M. & Scharschmidt, B. F. Electrophysiological evidence for Na^+ -coupled bicarbonate transport in cultured rat hepatocytes. *Am. J. Physiol. (Lond.)* **256**, G491–G500 (1989).
- Gleeson, D., Smith, N. D. & Boyer, J. L. Bicarbonate-dependent and -independent intracellular pH regulatory mechanisms in rat hepatocytes. *J. Clin. Invest.* **84**, 312–321 (1989).
- Weintraub, W. H. & Machen, T. E. pH regulation in hepatoma cells: roles of Na^+/H^+ exchange, $\text{Cl}^-/\text{HCO}_3^-$ exchange, and $\text{Na}^+/\text{HCO}_3^-$ cotransport. *Am. J. Physiol.* **257**, G317–G327 (1989).
- Ishiguro, H., Steward, M. C., Lindsay, A. R. G. & Case, R. M. Accumulation of intracellular HCO_3^- by $\text{Na}^+/\text{HCO}_3^-$ cotransport in interlobular ducts from the guinea-pig pancreas. *J. Physiol. (Lond.)* **495**, 169–178 (1996).
- Rajendran, V. M., Oesterlin, M. & Binder, H. J. Sodium uptake across basolateral membrane of rat distal colon. Evidence for Na^+/H^+ exchange and Na^+ -anion cotransport. *J. Clin. Invest.* **88**, 1379–1385 (1991).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, NY, 1989).
- Hediger, M. A., Coady, M. J., Ikeda, T. S. & Wright, E. M. Expression cloning and cDNA sequencing of the $\text{Na}^+/\text{glucose}$ co-transporter. *Nature* **330**, 379–381 (1987).
- Kanai, Y. & Hediger, M. A. Primary structure and functional characterization of a high-affinity glutamate transporter. *Nature* **360**, 467–471 (1992).
- Hediger, M. A. & Rhoads, D. B. Molecular physiology of sodium-glucose cotransporters. *Phys. Rev.* **74**, 993–1026 (1994).
- Reithmeier, R. A. F. The erythrocyte anion transporter (band 3). *Curr. Opin. Struct. Biol.* **3**, 515–523 (1993).
- Alper, S. L. The band 3-related anion exchanger (AE) gene family. *Annu. Rev. Physiol.* **53**, 549–564 (1991).

Acknowledgements. We thank M. O. Bevensee, B. A. Davis, P. Fong and N. L. Nakhoul for discussion on the manuscript; N. L. Nakhoul and S. A. Nakhoul for help with dissecting *Ambystoma* kidneys; and S. L. Alper, M. J. Jennings and R. A. F. Reithmeier for assistance with membrane structural assignments. This work was supported by grants from the NIH (M.A.H., E.L.B. and W.F.B.). M.E.R. was supported by the National Kidney Foundation and an individual NIDDK National Research Service Award.

Correspondence and requests for materials should be addressed to M.E.R. (e-mail: mromero@biomed.med.yale.edu, or mfr2@po.cwru.edu).

Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels

Karen F. Kozarsky*, Mary H. Donahee*, Attilio Rigotti†, Sohah N. Iqbal‡, Elazer R. Edelman§ & Monty Krieger†

* Institute for Human Gene Therapy and Department of Molecular and Cellular Engineering, University of Pennsylvania Medical Center, Philadelphia, Pennsylvania 19104, USA

† Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

§ Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Harvard-Massachusetts Institute of Technology, Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA

The risk of atherosclerosis, a leading cause of cardiovascular disease and death, is inversely related to plasma levels of high-density lipoprotein (HDL) cholesterol, although the mechanism of this protective effect is unclear¹. The class B scavenger receptor, SR-BI, is the first HDL receptor to be well defined at a molecular level and is a mediator of selective cholesterol uptake *in vitro*². It is expressed most abundantly in steroidogenic tissues, where it is coordinately regulated with steroidogenesis by adrenocorticotrophic hormone (ACTH), human chorionic gonadotropin (hCG) and oestrogen, and in the liver, where its expression in rats is suppressed by oestrogen^{3,4}. Here we show that adenovirus-mediated, hepatic overexpression of SR-BI in mice on both sinusoidal and canalicular surfaces of hepatocytes results in the virtual disappearance of plasma HDL and a substantial increase in biliary cholesterol. SR-BI may directly mediate these effects by increasing hepatic HDL cholesterol uptake or by increasing cholesterol secretion into bile, or both. These results indicate that SR-BI may be important in hepatic HDL metabolism, in determining plasma HDL concentrations, and in controlling cholesterol concentrations in bile, and thus may influence the development and progression of atherosclerosis and gallstone disease.

The *in vivo* effects of murine SR-BI (mSR-BI) on HDL and biliary cholesterol metabolism were studied in C57BL/6 mice that transiently overexpressed hepatic SR-BI after infection by intravenous infusion with a recombinant, replication-defective adenovirus (Ad.mSR-BI). In the Ad.mSR-BI virus, the mSR-BI complementary DNA is under the control of the cytomegalovirus (CMV) immediate-early enhancer/promoter. Controls included uninfected mice and mice infected with recombinant, replication-defective adenovirus which either express β -galactosidase (Ad.lacZ, ref. 5) or lack a cDNA transgene (Ad.E1 Δ ⁶). Following intravenous injection of

recombinant adenovirus in mice, > 99% of transgene expression is localized to the liver⁷. Liver tissue from uninfected mice and mice infected with Ad.lacZ or Ad.E1 Δ exhibited relatively low levels of SR-BI expression, as determined by immunohistochemistry and immunofluorescence microscopy (Fig. 1b, d, respectively) and by immunoblotting of whole-tissue lysates (Fig. 1g). Hepatic SR-BI expression was dramatically increased three days after infection with Ad.mSR-BI (Fig. 1a, c, g). Although the amounts of SR-BI protein decreased with time after infection, we routinely observed levels substantially above those of controls 21 days after infection (Fig. 1g, and data not shown). Immunohistochemical and immunofluorescence staining (Fig. 1a–d) revealed that virtually all of the increase in SR-BI expression was localized to the surface of the hepatocytes, including both the sinusoidal (arrows) and bile canalicular (for example, punctate staining; arrowheads) domains. The relative amounts of apical and sinusoidal expression have not yet been determined. The substantial expression of SR-BI on hepatic canalicular surfaces in Ad.mSR-BI infected animals was unexpected, because an HDL receptor would be expected to be expressed primarily on the sinusoidal surfaces that face the plasma. Additional studies will be required to determine whether the canalicular localization in the SR-BI-overexpressing animals reflects ectopic expression or the normal subcellular distribution of this protein.

The effects of hepatic SR-BI overexpression on plasma lipoproteins were determined using fast pressure liquid chromatography (FPLC) (Fig. 2). As previously reported (for example, see ref. 8), most of the cholesterol in uninfected mice was in the HDL fraction,

Table 1 Plasma cholesterol and apoA-I levels

Day (number of animals)	Cholesterol (mg dl ⁻¹)		apoA-I (mg dl ⁻¹)	
	Ad.E1 Δ	Ad.mSR-BI	Ad.E1 Δ	Ad.mSR-BI
0* (10)	131.0	117.8	33.2	32.6
3 (2)	125.5	16.5	31.0	5.0
7 (2)	146.0	173.0	33.5	23.4
14 (2)	129.0	152.0†	32.5	26.0
21 (2)	113.0	87.5	34.0	32.0

Mice were infused with recombinant adenovirus and, on the indicated days, plasma samples were analysed for total cholesterol and apoA-I content. Individual plasma cholesterol values were within 19% of the average value for day 0, and within 14% for all other days; individual apoA-I determinations were within 10% of the average value for each datum.

* Day 0 samples were obtained before infusion of recombinant adenovirus.

† Single determination only.

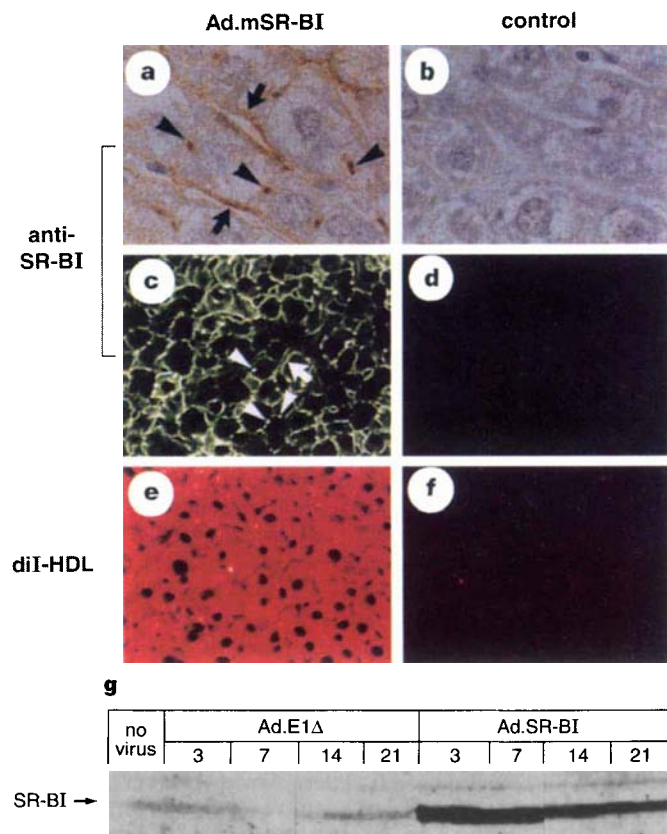


Figure 1 Effects of recombinant adenovirus infection on SR-BI expression in murine liver. **a–d**, Immunohistochemical (**a**, **b**) and immunofluorescence (**c**, **d**) analysis of liver tissues 3 or 5 days after injection of recombinant adenovirus. **a**, **c**, Ad.mSR-BI; **b**, control adenovirus Ad.lacZ; **d**, control adenovirus Ad.E1 Δ . **e**, **f**, Dil uptake from DiI-HDL in mice infected with Ad.mSR-BI (**e**) or Ad.E1 Δ (**f**). **c**, **e**, and **d**, **f** show the same fields photographed with either a fluorescein filter package (**c**, **d**) or rhodamine filter package (**e**, **f**). Samples were photographed at $\times 720$ (**a**, **b**) or $\times 200$ (**c**–**f**) magnification. **g**, Immunoblot analysis of SR-BI in lysates from livers collected at the indicated times after infection.

with small or undetectable amounts in the very-low-density lipoprotein (VLDL) and intermediate-density lipoprotein/low-density lipoprotein (IDL/LDL) fractions (Fig. 2a, 'pre'). Infection with the control Ad.E1 Δ virus had no effect on the lipoprotein profiles on day 3 (Fig. 2a) and relatively minor effects thereafter (Fig. 2c). Infection with Ad.mSR-BI dramatically reduced HDL cholesterol and lowered cholesterol in the IDL/LDL fractions by day 3 (Fig. 2b). Hepatic SR-BI levels in these animals gradually declined (Fig. 1g) and HDL cholesterol gradually increased (Fig. 2d). Quantitative analysis of the amounts of total cholesterol and apolipoprotein A-I (apoA-I), the major protein component of HDL in the plasma, were consistent with the FPLC analysis (Table 1). On days 7 and 10 after Ad.mSR-BI infection, when the hepatic SR-BI was still elevated and HDL cholesterol was significantly below control values, there was a substantial transient increase in both the VLDL and IDL/LDL cholesterol levels, which returned to near preinfection levels by day 21 (Fig. 2d). This increase in VLDL and IDL/LDL was not due to a decrease in hepatic LDL receptor expression in Ad.mSR-BI-infected animals relative to controls, as determined by immunoblot analysis (data not shown), and thus may in part be a consequence of altered lipoprotein secretion and metabolism in the plasma. The increased concentrations of cholesterol in the larger lipoproteins

seen on days 7 and 10 are reminiscent of the increases that characterize some forms of human combined hyperlipidaemia, suggesting that altered expression of SR-BI might be relevant in some cases of human dyslipidaemias. By day 21, the shape of the FPLC cholesterol profile was similar to that for control animals; however, the HDL cholesterol level had not returned to the preinfection value. This is presumably because there was still substantial overexpression of SR-BI in the liver. The numerous changes in the metabolism of most lipoproteins during the complex response to SR-BI overexpression may be related to the previously reported reciprocal changes in the plasma levels of large lipoproteins (such as chylomicron/VLDL remnants) and HDL (ref. 9).

Because the Ad.mSR-BI-mediated reduction in plasma HDL cholesterol and apoA-I could have been due to either reduced HDL synthesis, increased HDL clearance, or both, we examined the fate of labelled HDLs injected into mice 5 days after infection. The labelled HDLs included 125 I-HDL, in which the majority of the label was covalently bound to apoA-I, or HDL labelled with the fluorescent lipid DiI. We have previously used cellular DiI accumulation from DiI-HDL to assess SR-BI-mediated uptake of HDL lipid *in vitro*² and *in vivo*³. Plasma clearance of intravenously injected 125 I-HDL was significantly more rapid in Ad.mSR-BI animals (Fig. 2e,

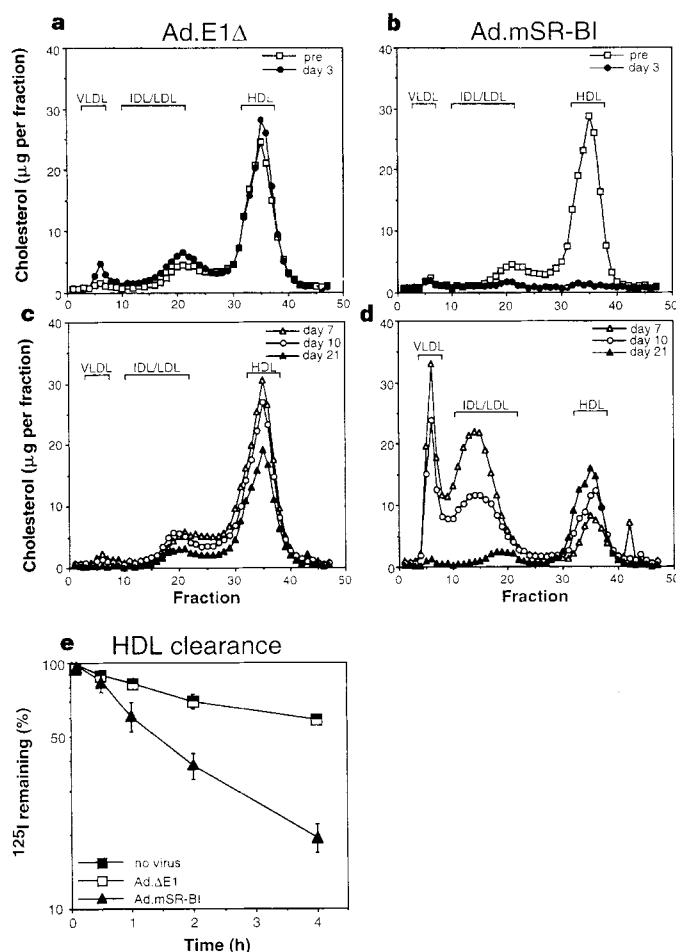


Figure 2 Lipoprotein cholesterol profiles and kinetics of 125 I-HDL clearance. **a-d**, FPLC analysis of plasma lipoprotein cholesterol. Before infection (pre) and on the indicated days after infection, murine plasma samples were collected and pooled (2 mice per specimen) and fractionated by FPLC, and the cholesterol content of each fraction determined. Positions of VLDL, IDL/LDL and HDL standards are indicated by brackets. The mean sizes of the IDL/LDL on days 7 and 10 were significantly larger than those in control animals or Ad.mSR-BI-infected animals on day 21. **a, c**, Control Ad.E1 Δ infection; **b, d**, Ad.mSR-BI infection. **e**, Kinetics of

125 I-HDL clearance. Five days after infection, a bolus of 125 I-HDL was injected intravenously into mice and the amounts of radioactivity remaining in the plasma (expressed as the average per cent $\pm$ standard deviation) were determined at the indicated times after injection. Control uninfected mice (black squares), Ad.E1 Δ -infected mice (white squares), Ad.mSR-BI-infected mice (black triangles). For each group, $n = 4$; because of the log scale, the small error bars for some of the data are obscured by the larger symbols.

triangles; $t_{1/2} \sim 1.5$ h) than in uninfected or Ad.E1 Δ -infected controls (squares; $t_{1/2} > 4$ h; see also ref. 10). The increased 125 I-HDL clearance was not necessarily due to hepatic internalization and degradation, because even though the bulk of HDL cholesterol in rats is cleared by the liver by selective uptake, the bulk of the HDL apolipoprotein is cleared by the kidney¹¹. Selective lipid uptake, in contrast to whole lipoprotein endocytic uptake¹², involves internalization of a lipoprotein's lipid, but not its apolipoprotein, components^{13–17}. In a complementary experiment, we injected DiI-HDL intravenously, and 2 h later collected the livers and used fluorescence microscopy to detect DiI uptake and SR-BI expression (immunofluorescence). The levels of SR-BI and DiI in the Ad.mSR-BI-infected animals (Fig. 1c, e) were much higher than in Ad.E1 Δ -infected controls (Fig. 1d, f). Although SR-BI was observed mainly at the cell surfaces (discrete peripheral staining pattern), the DiI staining was diffusely distributed throughout the cells. These results are consistent with SR-BI at the sinusoidal surfaces of hepatocytes having a direct effect on HDL clearance and hepatic lipid uptake. Note that SR-BI overexpression may have had indirect effects on these observations because of the reduced endogenous HDL pool in the Ad.mSR-BI-infected animals; this will be the subject of future investigations.

Previous studies have demonstrated that cholesterol from HDL is preferentially secreted into the bile relative to cholesterol from other lipoproteins¹⁸. To explore the possibility that increased delivery of HDL cholesterol to the liver in SR-BI-overexpressing animals might alter biliary cholesterol secretion, we measured the concentrations of three key biliary lipids four days after infection (Table 2). Although there were no notable differences in the bile salt and phospholipid concentrations, the cholesterol content of the hepatic bile in Ad.mSR-BI-infected mice was about twice that in the uninfected and the Ad.E1 Δ -infected controls. This is a substantial increase in biliary cholesterol. Pharmacological and dietary manipulations only increase biliary cholesterol in rats ~ 2 – 4 -fold^{19–21} and the gallbladder cholesterol levels in humans with cholelithiasis are only ~ 30 – 50% increased over that in gallstone-free individuals (reviewed in ref. 22). The reciprocal relationship between plasma HDL cholesterol and biliary cholesterol levels observed in control and mSR-BI-overexpressing mice has also been seen for plasma HDL cholesterol and biliary cholesterol saturation in humans²³. Furthermore, high-dose oestrogen treatment decreased both SR-BI expression³ and biliary cholesterol secretion²⁴ in rats. It seems likely that at least some of this increased biliary cholesterol in the Ad.mSR-BI-infected mice was derived from HDL, because in preliminary experiments we found that 2 h after DiI-HDL injection there was a 20-fold increase in DiI concentration in the bile of these mice relative to Ad.E1 Δ -infected mice (data not shown). In addition to the potential SR-BI-mediated increase in the selective uptake of plasma HDL cholesterol at hepatic sinusoidal surfaces, overexpression of SR-BI on the canalicular surfaces may have contributed to increased biliary cholesterol levels by participating directly in the secretion of cholesterol into the bile. Note that there have been reports of apoA-I immunoreactivity in bile²⁵, raising the possibility that apoA-I or a homologue (such as anionic polypeptide fraction²⁶) might interact with SR-BI in the bile canaliculus and facilitate cholesterol transport. Regardless of the mechanism, SR-BI-mediated biliary cholesterol secretion might profoundly influence

biliary cholesterol concentrations and thus potentially alter the propensity for gallstone formation²².

These studies establish that hepatic overexpression of SR-BI results in a large decrease in plasma HDL and a substantial increase in cholesterol secretion into the bile. Thus, in addition to previous *in vitro*² and *in vivo*³ findings that have strongly implicated SR-BI as an important receptor for HDL metabolism in steroidogenic tissues, there are now data that indicate a role for SR-BI in hepatic HDL metabolism *in vivo*, although it is important to note that these findings are based on non-physiological levels of expression. Changes in plasma HDL cholesterol and bile cholesterol due to modulation of hepatic SR-BI expression at more physiological levels remain to be determined. Even though there are species-dependent differences in lipoprotein metabolism, physiologically relevant alterations in hepatic SR-BI expression in humans may greatly influence steady-state plasma HDL concentrations. If so, SR-BI may play a role in reverse cholesterol transport and in determining relative risk for atherosclerotic disease¹. If SR-BI can mediate lipid efflux, it may do so in other tissues and possibly even allow cholesterol export from foam cells in artery walls. These findings may provide new directions for the study of lipoprotein and lipid metabolism and suggest that SR-BI may be an attractive target for therapeutic intervention in atherosclerosis and gallstone disease. □

Methods

Recombinant adenoviruses. The mSR-BI cDNA² was cloned into the plasmid pAd.CMV link²⁷, using the restriction enzymes *Hind*III and *Kpn*I. The new plasmid, designated pAd.CMVmSR-BI, was linearized with *Nhe*I and cotransfected into 293 cells with adenovirus type 5 sub360 DNA. Recombinant virus, Ad.mSR-BI, was plaque-purified, expanded and purified on CsCl gradients as previously described²⁸. The control adenovirus, Ad.E1 Δ (ref. 6), contains identical E1 and E3 deletions, but contains no transgene expression cassette; Ad.lacZ, previously designated Ad.CMV/lacZ, also contains the same E1 and E3 deletions, and has the *lacZ* gene under control of the CMV enhancer/promoter².

Animals/plasma analysis. C57BL/6 male mice (Taconic) over 8 weeks of age were used for all experiments. Mice were injected via the tail vein with 1×10^{11} particles of recombinant adenovirus in 0.1 ml of PBS. Blood from the retro-orbital venous plexus was collected into heparinized capillary tubes. The cholesterol and apoA-I content of whole plasma was analysed using a Cobra/Fara autoanalyser. FPLC analysis was performed on pooled plasma samples (two mice per datum) as previously described²⁸, except that cholesterol content was determined using the cholesterol kit from Wako.

Clearance experiments. Mice were injected via the tail vein with 125 I-HDL ($15 \mu\text{g}$ protein per mouse, $500 \text{ c.p.m. ng}^{-1}$) in 0.1 ml PBS. Blood was collected into heparinized capillary tubes at the indicated times, and radioactivity in $10 \mu\text{l}$ aliquots of plasma was determined using a gamma counter. The average radioactivity observed 3 min after injection was defined as 100% of injected radioactivity.

Hepatic uptake and bile secretion of DiI from DiI-HDL. Each animal was anaesthetized by intraperitoneal injection of a mixture of ketamine (140 mg kg^{-1}) and xylazine (14 mg kg^{-1}). The abdomen was opened, and the gallbladder duct and the distal region of the common bile duct were ligated. Then the common bile duct was incised and cannulated. The mice were immediately injected intravenously with $40 \mu\text{g}$ protein of HDL labelled with 1,1'-dioctadecyl-3,3,3',3'-tetramethyl indocarbocyanine perchloride (DiI-HDL²) in 0.1 ml PBS. Bile samples were collected for 2 h, while the body temperature was maintained by placing the animals under a heat lamp. Mice were then perfused with PBS to remove circulating DiI-HDL, followed by 3% paraformaldehyde in PBS, and finally PBS containing 15% sucrose and 10% optimal cutting temperature compound (OCT) (Miles). Liver tissues were then embedded in OCT, frozen and sectioned for microscopy and then fixed again with 3% paraformaldehyde/PBS for 10 min at 4°C , and stained with anti-mSR-BI antibody (see below) to permit simultaneous localization of DiI and SR-BI. Aliquots of bile ($50 \mu\text{l}$) were diluted to 1.5 ml with PBS and analysed by spectrofluorimetry (excitation at 550 nm, emission at 565 nm).

Immunohistochemistry, immunofluorescence and immunoblotting. For immunohistochemistry, paraffin-embedded sections⁵ were stained with anti-

Table 2 Biliary lipid concentrations

Virus	Cholesterol (mM)	Bile salts (mM)	Phospholipid (mM)
No virus	0.49 ± 0.14	21 ± 6	4.0 ± 1.0
Ad.E1 Δ	0.57 ± 0.13	23 ± 11	3.6 ± 1.2
Ad.mSR-BI	$1.15 \pm 0.36^*$	20 ± 6	4.7 ± 1.5

Four days after adenovirus infusion, bile was collected from individual mice, and each sample analysed for cholesterol, bile salt and phospholipid content. Values represent averages $\pm$ standard deviation. No virus, $n = 13$; Ad.E1 Δ , $n = 8$; Ad.mSR-BI, $n = 12$.

* $P < 0.0005$ compared to both no virus and Ad.E1 Δ controls.

mSR-BI as previously described⁴. Under these conditions the relatively low endogenous levels of hepatic SR-BI are not detected. For immunofluorescence, fresh-frozen tissues were sectioned and stained with an anti-mSR-BI anti-peptide antibody² as described⁵. For immunoblotting, lysates were prepared from liver tissues, and subjected to 10% sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis and western blotting as previously described²⁹. Paraffin-embedded sections of liver were stained with haematoxylin and eosin for assessment of liver histology; all samples from virus-infused animals showed transient inflammation and necrosis that was self-limiting (see for example ref. 27), and had abated by day 21.

Bile lipid analysis. Bile samples were collected as described above for 1–2 h and immediately frozen (–20 °C). Subsequently, bile acid content was determined using a kit from Sigma and cholesterol and phospholipid contents were measured as previously described³⁰.

Received 18 December 1996; accepted 14 March 1997.

- Gordon, D. J. & Rifkind, B. M. High-density lipoprotein: the clinical implications of recent studies. *N. Engl. J. Med.* **321**, 1311–1316 (1989).
- Acton, S. *et al.* Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* **271**, 518–520 (1996).
- Landschulz, K. T., Pathak, R. K., Rigotti, A., Krieger, M. & Hobbs, H. H. Regulation of scavenger receptor, class B, type I, a high density lipoprotein receptor, in liver and steroidogenic tissues of the rat. *J. Clin. Invest.* **98**, 984–995 (1996).
- Rigotti, A. *et al.* Regulation by adrenocorticotrophic hormone of the *in vivo* expression of scavenger receptor class B type I (SR-BI), a high density lipoprotein receptor, in steroidogenic cells of the murine adrenal gland. *J. Biol. Chem.* **271**, 33545–33549 (1996).
- Kozarsky, K. F. *et al.* *In vivo* correction of low density lipoprotein receptor deficiency in the Watanabe heritable hyperlipidemic rabbit with recombinant adenoviruses. *J. Biol. Chem.* **269**, 13695–13702 (1994).
- Engelhardt, J. F. *et al.* Direct gene transfer of human CFTR into human bronchial epithelia of xenografts with E1-deleted adenoviruses. *Nature Genet.* **4**, 27–34 (1993).
- Herz, J. & Gerard, R. D. Adenovirus-mediated transfer of low density lipoprotein receptor gene acutely accelerates cholesterol clearance in normal mice. *Proc. Natl Acad. Sci. USA* **90**, 2812–2816 (1993).
- Yokode, M. R., Hammer, R. E., Ishibashi, S., Brown, M. S. & Goldstein, L. Diet-induced hypercholesterolemia in mice: prevention by overexpression of LDL receptors. *Science* **250**, 1273–1275 (1990).
- Willnow, T. E., Sheng, Z., Ishibashi, S. & Herz, J. Inhibition of hepatic chylomicron remnant uptake by gene transfer of a receptor antagonist. *Science* **264**, 1471–1474 (1994).
- Ishibashi, S., Brown, M. S., Goldstein, J. L., Gerard, R. D., Hammer, R. E. & Herz, J. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *J. Clin. Invest.* **92**, 883–893 (1993).
- Glass, C., Pittman, R. C., Weinstein, D. B. & Steinberg, D. Dissociation of tissue uptake of cholesterol ester from that of apoprotein A-I of rat plasma high density lipoprotein: selective delivery of cholesterol ester to liver, adrenal, and gonad. *Proc. Natl Acad. Sci. USA* **80**, 5435–5439 (1983).
- Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).
- Pittman, R. C., Knecht, T. P., Rosenbaum, M. S. & Taylor, C. A. A nonendocytotic mechanism for the selective uptake of high density lipoprotein-associated cholesterol esters. *J. Biol. Chem.* **262**, 2443–2450 (1987).
- Leitersdorf, E., Stein, O., Eisenberg, S. & Stein, Y. Uptake of rat plasma HDL subfractions labeled with [3H]cholesteryl linoleyl ether or with 125I by cultured rat hepatocytes and adrenal cells. *Biochim. Biophys. Acta* **796**, 72–82 (1984).
- Gwynne, J. T. & Mahafee, D. D. Rat adrenal uptake and metabolism of high density lipoprotein cholesteryl esters. *J. Biol. Chem.* **264**, 8141–8150 (1989).
- Reaven, E., Chen, Y.-D. I., Spicher, M. & Azhar, S. Morphological evidence that high density lipoproteins are not internalized by steroid-producing cells during *in situ* organ perfusion. *J. Clin. Invest.* **74**, 1384–1397 (1984).
- Rinninger, F. *et al.* Selective uptake of high-density lipoprotein-associated cholesteryl esters by human hepatocytes in primary culture. *Hepatology* **19**, 1100–1114 (1994).
- Botham, K. M. & Bravo, E. The role of lipoprotein cholesterol in biliary steroid secretion: studies with *in vivo* experimental models. *Prog. Lipid Res.* **34**, 71–97 (1995).
- Turley, S. & Dietschy, J. M. In *The Liver, Biology and Pathobiology* (eds Arias, I. M., Jakoby, W. B., Popper, H., Schachter, D. & Schafritz, D. A.) 617–641 (Raven, New York, 1988).
- Rigotti, A., Marzolo, M. P. & Nervi, F. In *Current Topics in Membranes* (ed. Hoekstra, D.) 579–615 (Academic, San Diego, 1994).
- Rigotti, A., Marzolo, M. P., Ulloa, N. & Nervi, F. Effect of bean intake on biliary lipid secretion and on hepatic cholesterol metabolism in the rat. *J. Lipid Res.* **30**, 1041–1048 (1989).
- Carey, M. C. In *Bile Acids, Cholestasis and Gallstones* (eds H. Fromm, H. & Leuschner, U.) 147–175 (Kluwer, Dordrecht, 1996).
- Thornton, J. R., Heaton, K. W. & MacFarlane, D. G. A relation between high-density-lipoprotein cholesterol and bile cholesterol saturation. *Br. Med. J.* **283**, 1352–1354 (1981).
- Del Pozo, R., Nervi, F., Covarrubias, C. & Ronco, B. Reversal of progesterone-induced biliary cholesterol output by dietary cholesterol and ethynylestradiol. *Biochim. Biophys. Acta* **753**, 164–172 (1983).
- Sewell, R. B., Mao, S. J., Kawamoto, T. & LaRusso, N. F. Apolipoproteins of high, low, and very low density lipoproteins in human bile. *J. Lipid Res.* **24**, 391–401 (1983).
- Domingo, N. *et al.* Epitope mapping of the human biliary amphipathic, anionic polypeptide: similarity with a calcium-binding protein isolated from gallstones and bile, and immunologic cross-reactivity with apolipoprotein A-I. *J. Lipid Res.* **33**, 1419–1430 (1992).
- Ye, X. *et al.* Prolonged metabolic correction in adult ornithine transcarbamylase-deficient mice with adenoviral vectors. *J. Biol. Chem.* **271**, 3639–3646 (1996).
- Kozarsky, K. F., Jooss, K., Donahue, M., Strauss, J. F. & Wilson, J. M. Effective treatment of familial hypercholesterolemia in the mouse model using adenovirus-mediated transfer of the VLDL receptor gene. *Nature Genet.* **13**, 54–62 (1996).
- Kozarsky, K., Grossman, M. & Wilson, J. M. Adenovirus-mediated correction of the genetic defect in hepatocytes from patients with familial hypercholesterolemia. *Somat. Cell Mol. Genet.* **19**, 449–458 (1993).
- Ulloa, N., Garrido, J. & Nervi, F. Ultracentrifugal isolation of vesicular carriers of biliary cholesterol in native human and rat bile. *Hepatology* **7**, 234–244 (1987).

Acknowledgements. We thank J. Glick, D. Rader and J. Wilson for helpful discussions, and S. Xu, M. Penman, P. Smith and K. Hiester for technical assistance. This work was supported by grants from the National Institutes of Health National Heart, Lung and Blood Institute to M.K. and E.E., and the Whitaker Foundation in Biomedical Engineering, the Burroughs-Wellcome Fund in Experimental Therapeutics and the National Science Foundation to E.E. and a Howard Hughes Medical Institute Postdoctoral Fellowship to A.R.

Correspondence and requests for materials should be addressed to K.F.K. (e-mail: kozarsky@mail.med.upenn.edu).

Repression of the CDK activator Cdc25A and cell-cycle arrest by cytokine TGF- β in cells lacking the CDK inhibitor p15

Antonio Iavarone & Joan Massagué

Cell Biology and Genetics Program and Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The activity of the cyclin-dependent kinases (CDKs) that control cell growth and division can be negatively regulated by tyrosine phosphorylation or by the binding of various CDK inhibitors¹. Whereas regulation by tyrosine phosphorylation is well documented in CDKs that function during mitosis, little is known about its role in the regulation of CDKs that act in the G1 phase of the cell cycle². In contrast, much evidence has accumulated on the regulation of G1 CDKs by CDK inhibitors¹. The cytokine TGF- β inhibits growth by causing cell-cycle arrest as a result of increasing the concentration of the Cdk4/6 inhibitor p15^{INK4B/MTS2} (refs 3, 4). Here we report that TGF- β can also cause the inhibition of Cdk4 and Cdk6 by increasing their level of tyrosine phosphorylation. Tyrosine phosphorylation and inactivation of Cdk4/6 in a human mammary epithelial cell line are shown to result from the ability of TGF- β to repress expression of the CDK tyrosine phosphatase Cdc25A. Repression of Cdc25A and induction of p15 are independent effects mediating the inhibition of Cdk4/6 by TGF- β .

The protein p15 is an important mediator of the antiproliferative effects of TGF- β ^{3,4}. Its increased expression in response to TGF- β is sufficient to cause cell-cycle arrest in certain epithelial cells⁵. However, p15^{INK4B/MTS2} is mutated, deleted or transcriptionally silenced in various human malignancies and often also during cell immortalization *in vitro*^{6,7}. We found that TGF- β addition significantly inhibits DNA synthesis in several p15-defective human tumour cell lines (Table 1). Our survey also included the spontaneously immortalized human mammary epithelial cell line MCF10A. This cell line has a normal phenotype⁸ but lacks p15^{INK4B/MTS2} and p16^{INK4A/MTS1} (Fig. 1a, b) and so shows no p15 response to TGF- β (Fig. 1c). TGF- β has a strong antiproliferative effect on MCF10A cells which is similar to its effect on p15-positive keratinocytes and lung epithelial cells⁴. The TGF- β effect on MCF10A is potent (half-maximal inhibitory concentration, IC₅₀ = 2.7 pM TGF β 1; Fig. 1d) and leads to G1 arrest (Fig. 1e) and accumulation of the retinoblastoma protein (RB) in the hypophosphorylated, growth inhibitory state (Fig. 1f). Thus, p15 is not essential for growth inhibition by TGF- β in several cell types. Given their strong response and their otherwise normal phenotype, MCF-10A cells were chosen to investigate the p15-independent G1-arrest response to TGF- β .

The cyclin D-dependent kinases Cdk4 and Cdk6 have RB kinase activity and are essential for G1 progression^{9,10}. TGF- β addition to exponentially growing MCF10A cells caused a nearly complete loss

of cyclin D1- and Cdk6-associated RB kinase activities (Fig. 2a). This occurred without any loss of cyclin D1, cyclin D3, Cdk4 or Cdk6 (Fig. 2b) (cyclin D2 was not detected in MCF10A cells). Cdk2-associated kinase activity was also inhibited by TGF- β but this was associated with a loss of cyclin A (data not shown) which may be secondary to inhibition of Cdk4/6 (ref. 11). Immunoprecipitation of Cdk6 followed by immunoblot analysis of these precipitates with anti-cyclin D antibodies showed no loss in Cdk6-bound cyclins D1 or D3 in response to TGF- β (Fig. 2b). The levels of Cdk4-bound cyclins D1 and D3 were decreased (Fig. 2b) either because of a primary effect on the assembly or stability of these complexes, or because of a modification similar to that which inhibited, without dissociating, cyclin D-Cdk6 complexes. Inhibition of Cdk4 and Cdk6 could not be attributed to known CDK inhibitors as those

were absent (p15 and p16), not detectable (p57^{Kip2}), or expressed at very low levels (p18^{INK4C}, p19^{INK4D} and p27^{Kip1}) or at higher levels (p21^{Cip1}), but were not raised by TGF- β (data not shown). Collectively, these results indicate that the loss of cyclin D-dependent kinase activity in TGF β -treated MCF10A cells is not due to a loss of cyclin or CDK components, or to an increase in CDK inhibitors or, in the case of cyclin D-Cdk6, a dissociation of cyclin-CDK complexes.

CDKs can also be regulated by phosphorylation. The CDK-activating kinase CAK phosphorylates a threonine in the T-loop of various CDKs¹². The kinase Wee1 phosphorylates a tyrosine in the glycine loop of the mitotic CDK Cdc2, and the phosphatase Cdc25 reverses this effect². TGF- β treatment increases the overall phosphorylation of Cdk4, Cdk6 or Cdk2 only slightly, as deter-

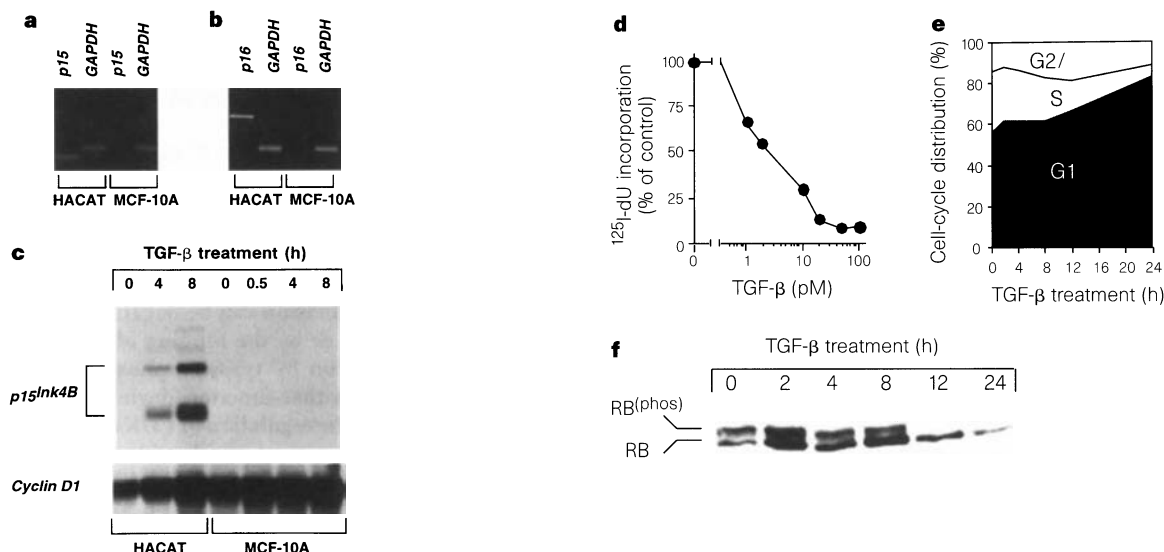


Figure 1 Properties of MCF10A human mammary epithelial cells. MCF10A cells have homozygous deletion of the *p15^{INK4B/MTS2}* and *p16^{INK4A/MTS1}* genes as determined by PCR amplification of exon 1 (**a**) or exon 2 (**b**), respectively. HaCaT human keratinocyte DNA was used as a positive control. A glyceraldehyde dehydrogenase (GAPDH) gene segment was amplified as an internal control. **c**, Northern blot analysis of *p15* mRNA and, as a control, cyclin D1

mRNA using poly(A)⁺ RNA from HaCaT and MCF-10A cells treated with TGF- β for the indicated times. **d**, Concentration-dependent inhibition of DNA synthesis by TGF- β addition (24 h) to exponentially growing MCF-10A cells. The data shown are the average of triplicate assays. **e**, Cell-cycle distribution, and **f**, RB protein western immunoblot assays of MCF10A cells treated with TGF- β for the indicated periods.

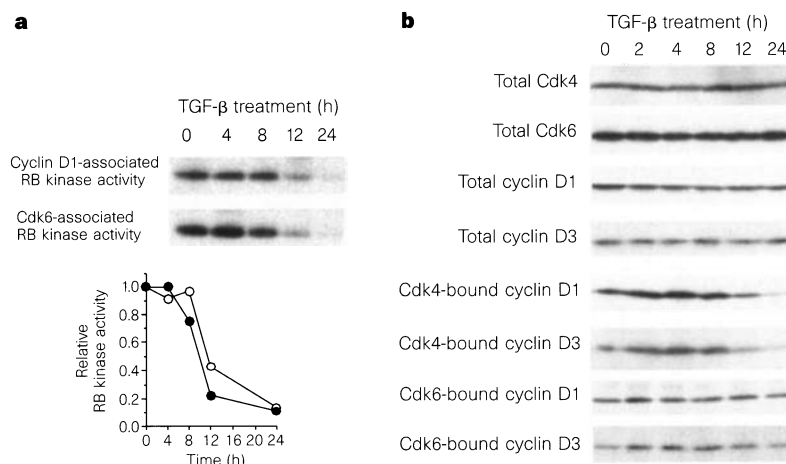


Figure 2 Inhibition of cyclin D-dependent kinases by TGF- β without changes in CDK components. **a**, MCF10A cells were treated with TGF- β for the indicated times. Cyclin D1 and Cdk6 were immunoprecipitated and the RB kinase activity of the immune complexes was assayed. The ³²P autoradiographic signal of the RB bands (top panels) was quantified and is plotted relative to the value in cells without TGF- β (white circles, cyclin D1; black circles, Cdk6). **b**, Total

amounts of the indicated proteins were determined by western immunoblotting of lysates from TGF β -treated MCF-10A cells (cyclins D1 and D3) or by immunoprecipitation followed by immunoblotting with the same antibody (Cdk4 and Cdk6). CDK-bound cyclin was determined by western immunoblotting of CDK immunoprecipitates.

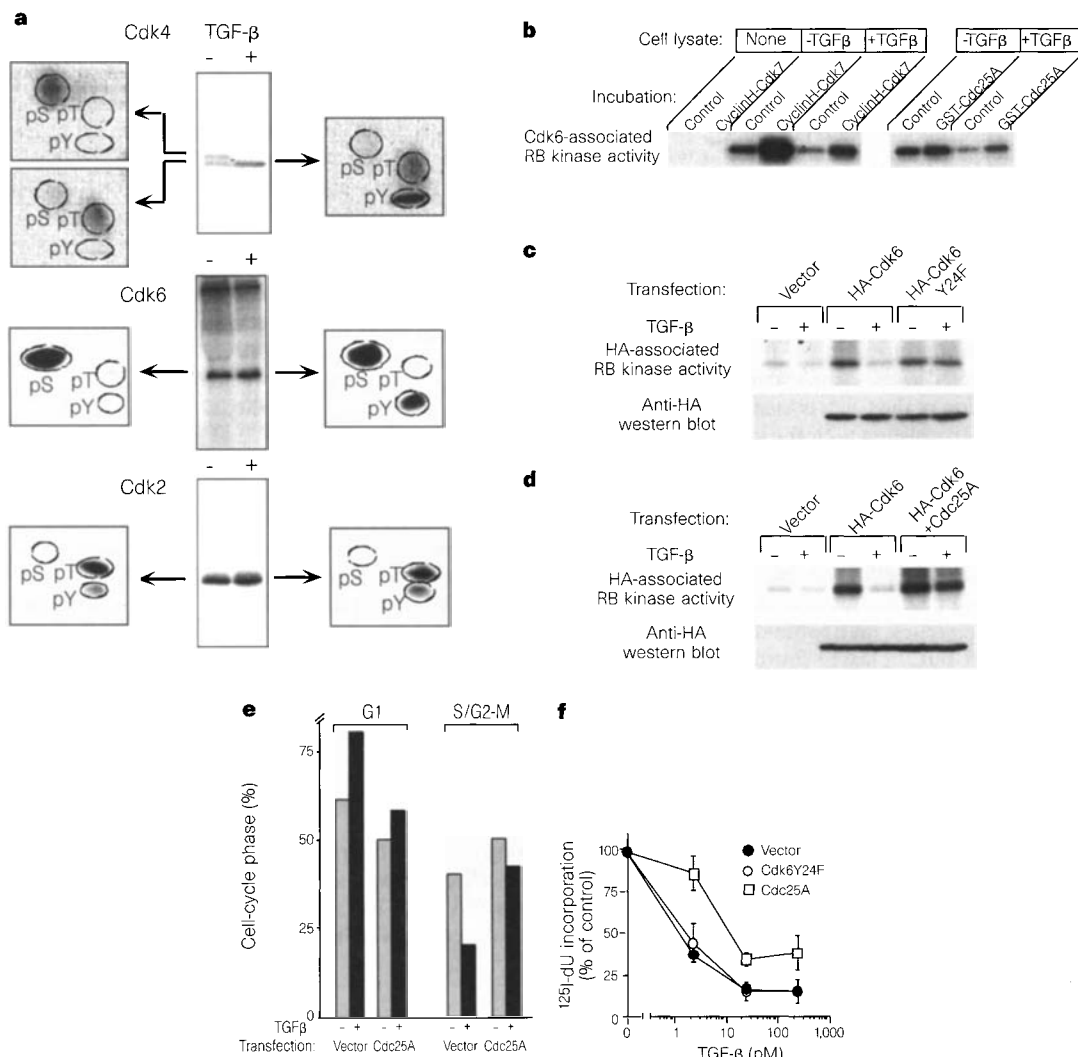


Figure 3 Cdk inhibition by TGF β -induced tyrosine phosphorylation. **a**, Cdk4, Cdk6 or Cdk2 were immunoprecipitated from 32 P-labelled MCF-10A treated with or without TGF β for 24 h and were subjected to acid hydrolysis and phosphoamino-acid analysis. The positions of phosphoserine (pS), phosphothreonine (pT) and phosphotyrosine (pY) are indicated. **b**, Effect of recombinant CAK (cyclin H-Cdk7) or GST-Cdc25A on the RB kinase activity of Cdk6 immunoprecipitated from control (– TGF β) or TGF β -treated (+TGF β) MCF-10A cells. The effect of CAK on a mock immunoprecipitation is also shown ('none'). MCF-10A cell extracts immunoprecipitated with normal rabbit serum contained no CAK-activated RB kinase activity (not shown). **c**, Cdk6(Y24F) is resistant to inhibition by TGF β . MCF-10A cells were transfected with pCMV vector encoding HA-tagged Cdk6 (HA-Cdk6), Cdk6 carrying a Y24F mutation (HA-Cdk6[Y24F]) or vector alone ('vector') and treated with TGF β for 24 h. RB kinase activity was determined in anti-HA immunoprecipitates. Expression levels of epitope-tagged proteins were determined by protein immunoblotting. **d**, MCF-10A cells transfected with pCMVNeobam 'vector', pCMV-Cdk6HA plus pCMV5 (HA-Cdk6), or with pCMV-Cdk6HA plus pCMVCdc25A (HA-Cdk6 + Cdc25A) were analysed as for **c**. **e**, Cells were cotransfected with the indicated vectors and green fluorescent protein vector, collected after treatment with TGF β for 24 h and sorted by fluorescence. DNA content was analysed by flow cytometry. The percentage of GFP-positive cells in G1 or S/G2-M is plotted. **f**, Cells were cotransfected with the indicated vectors plus green fluorescent protein vector and sorted. Fluorescent cells were replated and 6 h later incubated with TGF β as indicated. Data are the average $\pm$ s.d. of triplicate assays.

Table 1 Antiproliferative response to TGF β in cells with *p15^{INK4B}* homozygous deletions

Cell line	Cell type	125 I-dU incorporation (% of control)		TGF- β response
		5% FBS	0.5% FBS	
A427	Lung carcinoma	95	101	–
SK-LU-1	Lung carcinoma	102	119	–
SK-LC-16	Lung carcinoma	116	125	–
NCH-322	Lung carcinoma	82	93	–
A375*	Melanoma	104	129	–
SK-MEL-23	Melanoma	92	99	–
SK-MEL-44	Melanoma	104	100	–
SK-MEL-85	Melanoma	106	97	–
SK-MEL-93	Melanoma	107	96	–
SK-MEL-100	Melanoma	87	85	–
SK-MEL-245	Melanoma	105	88	–
U118	Glioma	111	142	–
SK-MG-10	Glioma	108	86	–
SK-MG-12	Glioma	108	101	–
BT-20	Breast carcinoma	73	76	$\pm$
SK-MEL-88	Melanoma	70	80	$\pm$
HepG2	Hepatoma	76	77	$\pm$
A549	Lung carcinoma	70	72	$\pm$
SK-BR-7	Breast carcinoma	58	63	+
MDA-MB-231	Breast carcinoma	58	61	+
SK-MEL-173*	Melanoma	57	63	+
MCF-10A	Mammary epithelium	10	ND	+

Percent [125 I]-iododeoxyuridine incorporation relative to untreated controls was determined after treatment with TGF β for 24 h. Assays were done with cells maintained in medium containing 0.5 or 5% of fetal bovine serum (FBS). MCF-10A cells did not grow in 0.5% serum. Values are the average of triplicate determinations. Standard deviations were never more than 10% of the average value. A decrease of less than 20% in the incorporation was not considered significant (–), and a decrease between 20 and 30% was considered marginal ($\pm$). All cell lines contain *p15^{INK4A/MTS1}* deletions except those marked with an asterisk (*). ND, not determined.

mined by CDK immunoprecipitation from ^{32}P -labelled MCF10A cells (Fig. 3a). However, phosphoamino-acid analysis revealed a marked increase in tyrosine phosphorylation of Cdk4 and Cdk6 in response to TGF- β (Fig. 3a). Cdk4 from control cells migrated as a tight doublet and Cdk6 as a single band, and both CDKs contained phosphoserine and phosphothreonine but almost no phosphotyrosine (Fig. 3a). After cell treatment with TGF- β , Cdk4 and Cdk6 migrated as single bands and showed a 10-fold increase in phosphotyrosine relative to untreated controls. Cdk2 showed only a minor increase in phosphotyrosine in response to TGF- β (Fig. 3a).

To determine the significance of this effect, we focused on Cdk6 because this kinase is inhibited by TGF- β treatment without losing bound cyclin D. Cdk6 immunoprecipitated from MCF10A cells gained activity when incubated with recombinant CAK¹² (Fig. 3b), suggesting that only a fraction of the Cdk6 present in these cells was in the active state. This was consistent with the low phosphothreonine content of Cdk6 (Fig. 3a). After incubation with CAK, the activity of Cdk6 from TGF β -treated cells was still lower than the activity of the same amount of Cdk6 from control cells (Fig. 3b). We also tested whether recombinant Cdc25 could restore activity to Cdk6 from TGF β -treated cells. Cdc25A was chosen because it is the isoform expressed during G1 phase¹³⁻¹⁵. Incubation with the fusion protein GST-Cdc25A increased less than twofold the kinase activity of Cdk6 from control cells but increased fourfold the activity of Cdk6 from TGF β -treated cells (Fig. 3b).

To demonstrate that Cdk6 inhibition by TGF- β is mediated by an increase in tyrosine phosphorylation, we introduced a Y24F mutation in the putative phosphorylation site¹⁶, transfected this construct into MCF10A cells, and selectively immunoprecipitated its product by using an epitope tag. Cdk6(Y24F) was not inhibited by TGF- β , whereas a wild-type Cdk6 was strongly inhibited (Fig. 3c). Furthermore, TGF- β inhibition of wild-type Cdk6 was prevented by cotransfection of Cdc25A (Fig. 3d), indicating that Cdc25A activity is rate-limiting in TGF β -treated cells. To confirm this, MCF10A

cells were transfected with a Cdc25A vector, incubated with TGF- β , and the transfectants sorted on the basis of the expression of a cotransfected GFP (green fluorescent protein) vector. Cdc25A transfection inhibited the TGF β -induced accumulation of cells in G1 (Fig. 3e; compare with Fig. 1e). Cdc25A overexpression also decreased TGF- β inhibition of RB phosphorylation, as determined by western immunoblotting of the sorted cell populations (data not shown). Furthermore, Cdc25A-transfected MCF10A cells that were incubated with TGF- β after sorting showed a significantly decreased growth inhibitory response compared to controls. Cdk6(Y24F) transfection did not have this effect (Fig. 3f), which is consistent with the idea that additional targets of Cdc25A (such as Cdk4) are required for cell proliferation.

In light of these results, we investigated whether TGF- β inhibits Cdc25A. TGF- β addition to MCF10A cells rapidly decreased *cdc25A* messenger (Fig. 4a, b). TGF- β induced a fivefold decrease in Cdc25A (Fig. 4c) which was parallel to the inhibition of cyclin D-dependent kinases (Fig. 2a). Cdc25A protein declined more slowly than its mRNA, possibly because the protein had a longer half-life. Cdc25A was also downregulated by TGF- β in primary human keratinocytes and immortalized HaCaT keratinocytes (data not shown), indicating that this effect can coexist with TGF β -induced upregulation of p15. Among tumour cell lines that are growth-inhibited by TGF- β , SK-BR-7 breast carcinoma cells showed Cdc25A downregulation, whereas A549 and BT-20 cells did not. Of the cell lines that are not growth-inhibited by TGF- β , those tested (SK-LC-16, A375 and U118) showed no Cdc25A downregulation.

Evidence that Cdc25A downregulation is a specific response to TGF- β and not an adaptive response to quiescence was obtained using mitogen-deprived MCF10A cells that were stimulated to enter the cell cycle by mitogen readdition. Cdc25A levels in mitogen-deprived MCF10A cells were nearly as high as in exponentially growing cells, were not altered by readdition of mitogens, and were

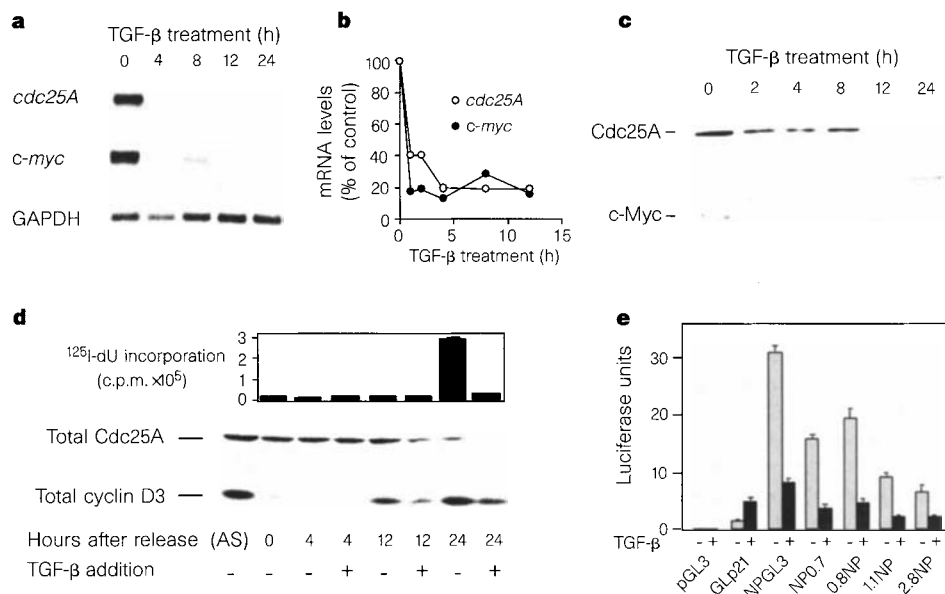


Figure 4 Cdc25A downregulation by TGF- β in MCF10A cells. **a**, Amounts of *cdc25A*, *c-myc* and *GAPDH* control mRNAs were determined by northern blotting. **b**, In a similar experiment using shorter times, mRNA was quantified with a phosphorimager and plotted relative to untreated controls. **c**, Western immunoblot assays of Cdc25A and c-Myc in lysates from TGF β -treated MCF10A cells. **d**, Cdc25A and cyclin D3 levels in MCF10A cells asynchronously growing (AS), cells made quiescent by mitogen deprivation (zero), or cells released from quiescence by readdition of mitogens for the indicated times. TGF- β was added to the indicated cultures at the time of mitogen readdition. Cell entry

into S phase was monitored by measuring ^{125}I -iododeoxyuridine incorporation over 2 h before each time point. **e**, Effect of TGF- β on the transcriptional activity of *cdc25A* promoter constructs controlling a luciferase reporter gene in transiently transfected HaCaT cells. Constructs containing the -400 to +108 promoter region of *cdc25A* (NPGL3) or this region plus a Myc-responsive region (NP0.7) or a Myc-unresponsive region (constructs 0.8NP, 1.1NP and 2.8NP) from the same gene have been described¹⁸. A p21 promoter construct (GLp21) which is activated by TGF- β ⁺ was used as a positive control.

downregulated by TGF- β (Fig. 4d). In comparison, cyclin D3 was markedly decreased by mitogen deprivation, was restored by readition of mitogens, as in other cell types⁹, but was not affected by TGF- β (Fig. 4d).

The kinetics of *cdc25A* downregulation were similar to those of *c-myc* downregulation (Fig. 4a, b), the latter being a TGF- β response described in many cell types¹⁷. *c-Myc* protein, which has a short half-life, disappeared in response to TGF- β with a $t_{1/2}$ = 2 h (Fig. 4c). Using a luciferase reporter gene under the control of the human *cdc25A* promoter region between nucleotides -400 and +108 relative to the first nucleotide of the *cdc25A* cDNA sequence¹⁸, we showed that TGF- β addition represses the activity of this promoter in HaCaT keratinocytes (Fig. 4e). Although *c-Myc* can act as a transcriptional activator of *cdc25A*, and a *c-Myc* response element has been identified within the second intron of the human *cdc25A* gene, no such element has been found in the -400 to +108 promoter region¹⁸. These results indicate that Cdc25A downregulation by TGF- β occurs at transcription. It remains to be determined whether *c-Myc* participates in this process.

CDK inhibition by TGF- β has been extensively investigated as a model antimitogenic response. Induction of p15 by TGF- β ^{3,4} is an important mediating step in this response^{4,5}. We have shown here that TGF- β can also inhibit Cdk4/6 by downregulating Cdc25A in a mammary epithelial cell line. Cdc25 is downregulated in other human epithelial cells that have or lack *p15*. A p15-independent mechanism must also mediate antiproliferative responses in non-epithelial cells because embryo fibroblasts from *p15*-null mice are growth inhibited by TGF- β (E. Latres, A.I. and M. Barbacid, unpublished results). Our results do not exclude the participation of a tyrosine kinase in CDK inhibition by TGF- β . Cdk4 downregulation¹⁹ and the CDK inhibitor p27^{Kip1} (ref. 20) are also implicated in the TGF- β response. However, Cdk4 is downregulated by TGF- β in lung epithelial cells when leaving G0 (ref. 19) but not when exponentially growing⁴. p27 levels are not affected by TGF- β ²¹ but are redistributed from cyclin D- to cyclin E-dependent kinases, inhibiting the latter, secondarily to an increase in p15 (refs 4, 5). Induction of p15 and downregulation of Cdc25A by TGF- β therefore constitute two complementary mechanisms of inhibition of cyclin D-dependent kinase. Overexpression of Cdc25 has been described in breast cancer²². Thus, defects in TGF- β antiproliferative responsiveness might result not only from losses in TGF- β receptors²³, signal transducers^{24,25} or p15, but also from a gain in Cdc25A activity. □

Methods

MCF10A cells were routinely grown in DMEM containing 5% horse serum, 10 μ g ml⁻¹ insulin, 50 μ g ml⁻¹ hydrocortisone and 20 ng ml⁻¹ epidermal growth factor. Growth inhibition by TGF- β (100 pM, unless noted otherwise) was evaluated from the incorporation of [¹²⁵I]iododeoxyuridine into DNA during the last 2 h of incubation. To achieve entry into and exit from quiescence, MCF-10A cells were incubated for 48 h in DMEM without supplements and then released from mitogen deprivation by addition of complete medium⁸. Flow cytometry analysis was done as described²⁰.

To study the effect of Cdc25A on TGF β -induced cell cycle arrest, MCF-10A cells were cotransfected with pGFP3 (Clontech) expressing the green fluorescent protein (GFP) plus a CMV-based vector (pcDNA3; Invitrogen) or pCMVCdc25A. Cells were treated with TGF- β and 2 $\times$ 10⁴ cells that were positive for green fluorescence were sorted and nuclear DNA content²⁶ and Rb phosphorylation were analysed. In another experiment, the sorted GFP-positive cells were cultured in different concentrations of TGF- β for 24 h and the incorporation of [¹²⁵I]iododeoxyuridine into DNA was determined.

HaCaT cells were transfected by the DEAE-dextran method and luciferase activity was measured in triplicate after cell incubation with or without TGF- β for 24 h. All DNA constructs were in pGL3 vector (Promega), except for the p21 promoter construct which was in pGL2 (Promega). Ten times more cell lysate was used in the luciferase assays of the p21 promoter construct.

The vector pCMVCdk6-HA has been described²⁷. A point mutation chan-

ging amino acid 24 from tyrosine to phenylalanine was introduced by PCR. DNA sequencing was used to confirm the correct sequence of the entire open reading frame. MCF10A cells were transfected with the corresponding HA-tagged constructs by a modified version of the DEAE-dextran method. Cells were treated with TGF- β between 24 and 48 h after transfection. Cell lysates were immunoprecipitated with the anti-HA monoclonal antibody 12CA5 (Zymed) and analysed for their associated Rb kinase activity.

Genomic DNA (100 μ g) from MCF-10A⁸ or HaCaT²⁸ was PCR-amplified with appropriate primers (primer sequences are available from the authors upon request) that yield a 132-bp *p15* fragment, a 510-bp *p16* fragment or a 190-bp *GAPDH* fragment. Northern blot analysis was done as described⁴. The probe for Cdc25A corresponded to a fragment (nucleotides 1,878–2,025 of human Cdc25A²⁹) that had been labelled to high specificity by PCR.

Cell lysates were prepared in 50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM EDTA, 2.5 mM EGTA, 10% glycerol, 1 mM DTT, 0.1% Tween 20, 10 mM β -glycerophosphate, 1 mM NaF, 0.1 mM sodium orthovanadate, 1 mM PMSF, 1 μ g ml⁻¹ aprotinin, 1 μ g ml⁻¹ leupeptin. Antibodies were against Rb (G3-245, Pharmingen), cyclin D1 (Upstate Biotechnology antibody for immunoblot; DCS-11 antibody³⁰ for kinase assays), cyclin D3 (Pharmingen), Cdk2 (Pharmingen), Cdk4 (Pharmingen and Santa Cruz Biotechnology), Cdk6 (ref. 10), Cdc25A (Santa Cruz Biotechnology) or *c-Myc* (Oncogene Science). Cell lysates were normalized based on protein content (Biorad). Immunoprecipitates were collected using protein A/G–Sephacrose (Santa Cruz Biotechnology). Immune complex kinase assays were done essentially as described^{9,30}. Standard immunoprecipitation with Cdk6 or cyclin D1 antibodies was followed by four washes in lysis buffer and two washes in kinase buffer before testing the kinase activity towards a C-terminal fragment of Rb.

The effect of CAK or Cdc25A on Cdk6 immunoprecipitates was tested *in vitro* with purified preparations of baculovirally expressed CAK¹² or with bacterially expressed GST–Cdc25A fusion protein¹³, respectively. Incubations were for 30 min at 30 °C before Rb kinase assays. For phosphoamino-acid analysis of CDKs, MCF-10A cells were labelled with 2.5 mCi [³²P]orthophosphate per ml for 5 h in phosphate-free DMEM containing 5% dialysed FBS and the required growth factors. Cells were lysed with RIPA buffer containing 10 mM NaF, 5 mM β -glycerophosphate and 2 mM sodium orthovanadate which was also included in the washing buffer. Phosphoamino-acid content of [³²P]-labelled Cdk6 was determined by partial hydrolysis in 6 M HCl and two-dimensional electrophoresis using the HTLE-7000 system (CBS Scientific).

Received 28 August 1996; accepted 28 March 1997.

- Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
- Morgan, D. O. Principles of CDK regulation. *Nature* **374**, 131–134 (1995).
- Hannon, G. J. & Beach, D. p15^{INK4B} is a potential effector of TGF β -induced cell-cycle arrest. *Nature* **371**, 257–261 (1994).
- Reynisdóttir, L., Polyak, K., Iavarone, A. & Massagué, J. Kip/Cip and Ink4 Cdk inhibitors cooperate to induce cell cycle arrest in response to TGF- β . *Genes Dev.* **9**, 1831–1845 (1995).
- Reynisdóttir, L. & Massagué, J. The subcellular locations of p15^{INK4B} and p27^{Kip1} coordinate their inhibitory interactions with cdk4 and cdk2. *Genes Dev.* **11**, 492–503 (1997).
- Kamb, A. *et al.* A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264**, 436–440 (1994).
- Nobori, T. *et al.* Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. *Nature* **368**, 753–756 (1994).
- Soule, H. D. *et al.* Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* **50**, 6075–6086 (1990).
- Matsushime, H. *et al.* D-type cyclin-dependent kinase activity in mammalian cells. *Mol. Cell. Biol.* **14**, 2066–2076 (1994).
- Meyerson, M. & Harlow, E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol. Cell. Biol.* **14**, 2077–2086 (1994).
- Schulze, A. *et al.* Cell cycle regulation of the cyclin A gene promoter is mediated by a variant E2F site. *Proc. Natl Acad. Sci. USA* **92**, 11264–11268 (1995).
- Fisher, R. P. & Morgan, D. O. A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* **78**, 713–724 (1994).
- Hoffmann, J., Draetta, G. & Karsenti, E. Activation of the phosphatase activity of human cdc25A by a cdk2-cyclin E dependent phosphorylation at the G1/S transition. *EMBO J.* **13**, 4302–4310 (1994).
- Jinno, S. *et al.* Cdc25A is a novel phosphatase functioning early in the cell cycle. *EMBO J.* **13**, 1549–1556 (1994).
- Terada, Y., Tatsuka, M., Jinno, S. & Okayama, H. Requirement for tyrosine phosphorylation of Cdk4 in G1 arrest induced by ultraviolet irradiation. *Nature* **376**, 358–362 (1995).
- Meyerson, M. *et al.* A family of human cdc2-related protein kinases. *EMBO J.* **11**, 2909–2917 (1992).
- Pietenpol, J. A. *et al.* TGF- β 1 inhibition of *c-myc* transcription and growth in keratinocytes is abrogated by viral transforming protein with pRB binding domains. *Cell* **61**, 777–785 (1990).
- Galaktionov, K., Chen, X. & Beach, D. Cdc25 cell-cycle phosphatase as a target of *c-Myc*. *Nature* **382**, 511–517 (1996).
- Ewen, M. E., Sluss, H. K., Whitehouse, L. L. & Livingston, D. M. TGF- β inhibition of Cdk4 synthesis is linked to cell cycle arrest. *Cell* **74**, 1009–1020 (1993).
- Polyak, K. *et al.* p27^{Kip1}, a cyclin-Cdk inhibitor, links transforming growth factor- β and contact inhibition to cell cycle arrest. *Genes Dev.* **8**, 9–22 (1994).

21. Polyak, K. *et al.* Cloning of p27Kip1 a cyclin-cdk inhibitor and a potential mediator of extracellular antimitogenic signals. *Cell* **78**, 59–66 (1994).
22. Galaktionov, K. *et al.* CDC25 phosphatases as potential human oncogenes. *Science* **269**, 1575–1577 (1995).
23. Markowitz, S. *et al.* Inactivation of the type II TGF β receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336–1338 (1995).
24. Eppert, K. *et al.* MADR2 maps to 18q21 and encodes a TGF β -regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543–552 (1996).
25. Hahn, S. A. *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).
26. Giaretti, W. & Nüsse, M. Light scatter of isolated cell nuclei as a parameter discriminating the cell-cycle subcompartments. *Meth. Cell Biol.* **41**, 389–400 (1994).
27. van den Heuvel, S. & Harlow, E. Distinct role for cyclin-dependent kinases in cell cycle control. *Science* **262**, 2050–2054 (1993).
28. Boukamp, P. *et al.* Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.* **106**, 761–771 (1988).
29. Galaktionov, K. & Beach, D. Specific activation of cdc25 tyrosine phosphatases by B-type cyclins: evidence for multiple roles of mitotic cyclins. *Cell* **67**, 1181–1194 (1991).
30. Lukas, J., Bartokova, J., Rohde, M., Strauss, M. & Bartek, J. Cyclin D1 is dispensable for G1 control in retinoblastoma gene-deficient cells independently of cdk4 activity. *Mol. Cell Biol.* **15**, 2600–2611 (1995).

Acknowledgements. We thank D. Beach, S. Brooks, H. Chou, K. Galaktionov, E. Harlow, A. Houghton, M. Meyerson, D. Morgan, M. Pagano and A. Russo for reagents, and T. Delohery for sorting and flow cytometry assays. This work was supported by NIH Breast Spore and Cancer Center grants. A.I. is supported in part by funds from the C. H. Revson and C. A. Dana foundations. J.M. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to J.M. (e-mail: j-massague@ski.mskcc.org).

Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F

Gustavo Leone*, James DeGregori*, Rosalie Sears, Laszlo Jakoi & Joseph R. Nevins

Department of Genetics, Howard Hughes Medical Institute, Duke University Medical Center, Box 3054, Durham, North Carolina 27710, USA

* These authors contributed equally to this work.

Considerable evidence points to a role for G1 cyclin-dependent kinase (CDK) in allowing the accumulation of E2F transcription factor activity and induction of the S phase of the cell cycle^{1,2}. Numerous experiments have also demonstrated a critical role for both Myc and Ras activities in allowing cell-cycle progression³. Here we show that inhibition of Ras activity blocks the normal growth-dependent activation of G1 CDK, prevents activation of the target genes of E2F, and results in cell-cycle arrest in G1. We also show that Ras is essential for entry into the S phase in *Rb*^{+/+} fibroblasts but not in *Rb*^{-/-} fibroblasts, establishing a link between Ras and the G1 CDK/Rb/E2F pathway. However, although expression of Ras alone will not induce G1 CDK activity or S phase, coexpression of Ras with Myc allows the generation of cyclin E-dependent kinase activity and the induction of S phase, coincident with the loss of the p27 cyclin-dependent kinase inhibitor (CKI). These results suggest that Ras, along with the activation of additional pathways, is required for the generation of G1 CDK activity, and that activation of cyclin E-dependent kinase in particular depends on the cooperative action of Ras and Myc.

To explore possible connections in the regulatory pathways involving Ras, Myc and G1 CDKs, we have generated a recombinant adenovirus that expresses a Ras dominant-negative mutant protein (Ras^{N17}). As shown by the experiment in Fig. 1a, when quiescent REF52 cells were stimulated by serum addition, inhibition of Ras activity through the expression of the Ras^{N17} mutant blocked the induction of S phase, consistent with previous studies^{4–6}. Moreover, inhibition of Ras also blocked the activation of the cyclin D- and cyclin E-dependent kinases that is normally observed following serum stimulation (data not shown) and prevented the appearance of the hyperphosphorylated form of Rb that normally occurs during the G1/S transition (Fig. 1b). Given that Ras is required for the

induction of G1 CDK activity and the phosphorylation of Rb, we have also examined the effect of Ras^{N17} expression on the accumulation of E2F activity. As seen by the northern analysis shown in Fig. 1c, the induction of several known E2F target genes was blocked in cells expressing the Ras^{N17} mutant.

The experiment shown in Fig. 1d suggests that Rb may be a critical target for Ras in the induction of S phase. The Ras^{N17} mutant was expressed in either normal or Rb-deficient fibroblasts derived from *Rb*^{-/-} mice, and S-phase entry was measured by

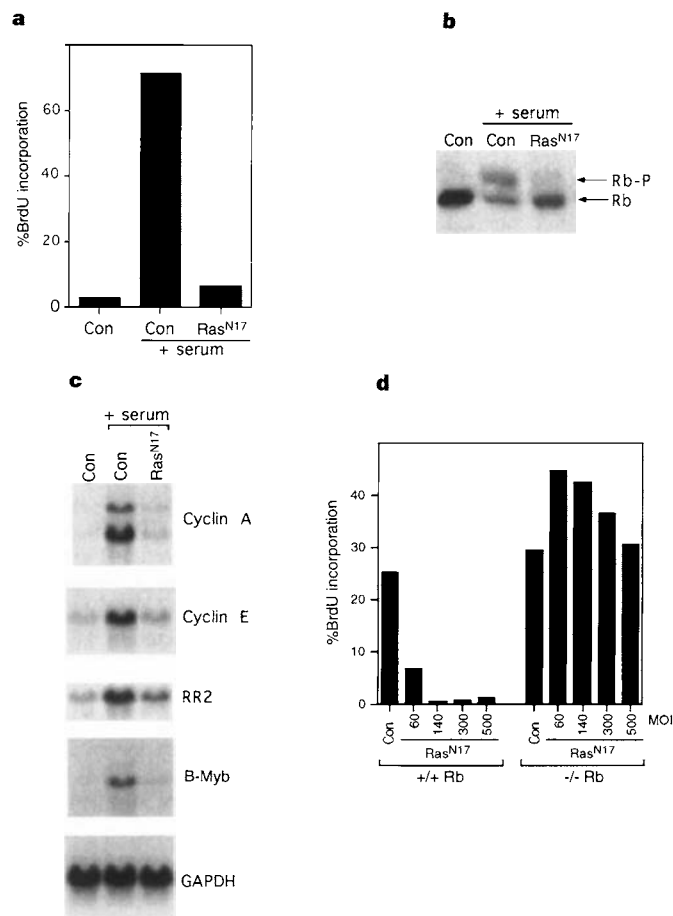


Figure 1 Ras is required for the generation of G1 CDK activity and E2F activity during the induction of S phase. **a**, Inhibition of S-phase induction of Ras^{N17}. REF52 cells were grown in D-MEM containing 0.25% serum for 48 h and then infected with either Ad-Ras^{N17} or Ad-Con (control virus with empty expression cassette) at a multiplicity of infection (MOI) of 100 FFU per cell. Where indicated, the cells were then stimulated by replacement with media containing 10% serum (+ serum) or were left in low-serum conditions. Cells were labelled with BrdU from 12 to 24 h post-serum addition and BrdU incorporation was detected by indirect immunofluorescence as previously described²⁰. At least 300 nuclei were scored for BrdU immunofluorescence. **b**, Inhibition of Ras activity prevents phosphorylation of Rb. Cells treated as in **a** were collected 18 h after serum addition in SDS loading buffer and Rb was detected by western blotting as described elsewhere²⁴ using anti-RB monoclonal antibody 14001A (Pharmingen). **c**, Inhibition of Ras activity blocks activation of E2F target genes. Cells treated as in **a** were collected 18 h after serum addition and processed for northern analysis. Probes for cyclin A, cyclin E, ribonucleotide reductase-2 (RR2), B-Myb, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) have been described previously²⁰. **d**, Ras is not required for the induction of S phase in *Rb*^{-/-} mouse fibroblasts. Early passage primary fibroblasts from *Rb* mutant or wild-type mouse embryos at 12.5 days gestation²⁵ were rendered quiescent by a 72-h incubation in D-MEM containing 0.2% FBS. Cells were infected with Ad-Ras^{N17} and subsequently stimulated with serum as in **a** and analysed for BrdU incorporation.

bromodeoxyuridine (BrdU) incorporation. As found for the REF52 cells, expression of Ras^{N17} blocked the induction of S phase in *Rb*^{+/+} mouse embryo fibroblasts (MEF) (Fig. 1d). In sharp contrast, Ras activity was not required for S-phase induction in *Rb*^{-/-} MEFs. Taken together, these results suggest that the Ras pathway plays an essential role in the G1 CDK/Rb pathway leading to the accumulation of E2F activity and the induction of DNA replication.

To analyse further the role of Ras in the G1 CDK/Rb/E2F pathway, we tested whether the overexpression of activities suspected to be downstream from the action of Ras would overcome the block resulting from the action of the Ras^{N17} mutant. Indeed, expression of cyclin E and Cdk2 bypassed the Ras^{N17}-mediated inhibition of DNA replication, allowing cells to proceed into and through S phase (Fig. 2a). The expression of cyclin E/Cdk2, which results in the phosphorylation of Rb (data not shown), also led to the restoration of E2F activity (Fig. 2b), as demonstrated by the activation of known E2F target genes. We also tested the ability of E2F overexpression to bypass the block imposed by the Ras^{N17} mutant. As shown in Fig. 2, expression of the E2F2 protein bypassed the Ras^{N17} block, allowing activation of E2F target genes and the induction of DNA replication. On the basis of these results, we conclude that both E2F and cyclin E/Cdk2 are important downstream effectors of Ras.

Given the requirement for Ras activity in allowing the induction of G1 CDK activity, Rb phosphorylation, E2F accumulation and DNA replication, we investigated whether the production of Ras activity would be sufficient to induce kinase activity and S-phase entry. Ras activity was introduced by infecting quiescent cells with a recombinant adenovirus containing the activated H-Ras^{61L} gene, and S-phase induction was scored by BrdU incorporation. Expression of the *Ras* gene alone was not sufficient to induce S phase under the conditions of these experiments (Fig. 3a). Because various experiments have demonstrated a collaboration between c-Myc (Myc) and Ras in transformation assays^{7,8}, we have assessed a possible interaction between these two activities by coexpressing Myc and Ras through recombinant viruses. Although more active

than Ras, Myc alone was also ineffective at inducing S phase. In contrast, the coexpression of both Myc and Ras led to a marked stimulation of DNA replication. The ability of Myc and Ras to induce S phase was dependent on the generation of G1 CDK activity as introduction of the p21 CKI protein into these cells blocked the Myc/Ras collaborative effect. Indeed, assays for kinase activity in cells expressing Myc and Ras revealed an induction of cyclin E/Cdk2 kinase activity, whereas Myc or Ras alone did not stimulate this activity (Fig. 3b). Surprisingly, cyclin D1-dependent kinase activity was not generated in these cells, and both the Rb and p107 proteins remained in the hypophosphorylated state (Fig. 3b,c). This suggests that, although overexpression of cyclin E/Cdk2 can result in the phosphorylation of Rb⁹, under normal circumstances Rb phosphorylation requires the prior action of cyclin D-dependent kinase, a conclusion reached in several other recent studies^{10,11}.

Two observations suggested that the synergistic action of Myc and Ras was not due to a simple augmentation of the level of either of these activities. Expression of Ras in quiescent fibroblasts resulted in the phosphorylation of Erk1 and Erk2, as indicated by a shift in their mobility in SDS gels or as detected by phospho-specific antibodies, and this effect was not enhanced by coexpression of Myc (Fig. 3d). Conversely, the ability of Myc to activate the expression of the *cdc25A* gene, recently shown to be a transcriptional target for the Myc protein¹², was not significantly affected by the coexpression of Ras (Fig. 3d). Therefore, the synergy observed between Myc and Ras in the stimulation of S phase and induction of cyclin E/Cdk2 activity seems to represent the cooperative action of the two proteins.

The fact that Myc and Ras expression led to S phase without Rb phosphorylation was surprising in light of the role of unphosphorylated Rb in controlling E2F and the substantial evidence linking the accumulation of E2F activity with S-phase induction. However, as shown in Fig. 3e, expression of Myc in quiescent cells led to the activation of the *E2F1* gene. The activation of the *E2F1* gene was not merely a consequence of cell-cycle progression, as expression of the p21 CKI, which blocked G1/S progression, did not affect the activation. Although it remains to be shown that the

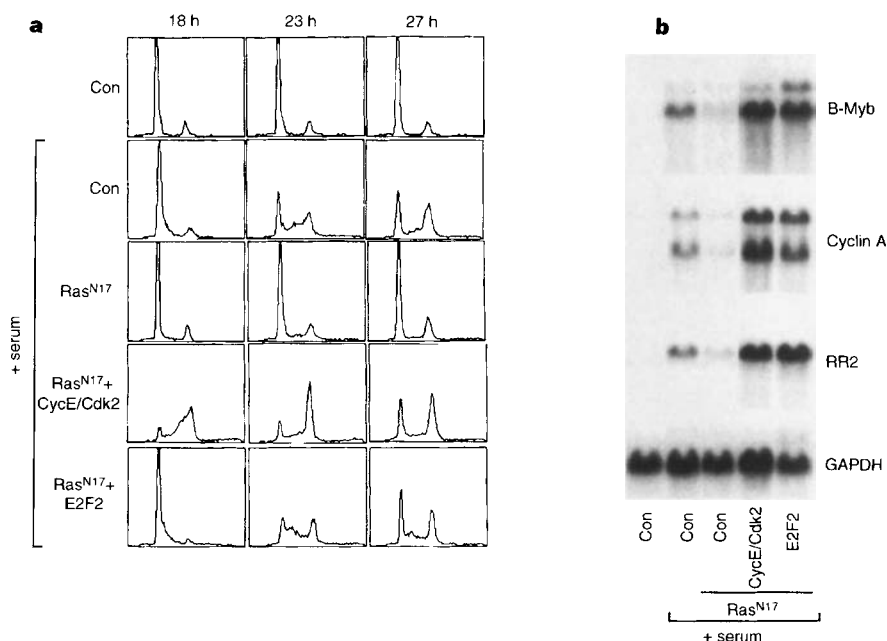


Figure 2 Either E2F or cyclin E/Cdk2 overexpression can bypass G1 arrest due to Ras inhibition. **a**, REF52 cells were plated, brought to quiescence and infected as in Fig. 1a with Ad-con (MOI 100), Ad-Ras^{N17} (MOI 100), Ad-CycE (MOI 200), Ad-Cdk2 (MOI 200) and Ad-E2F2 (MOI 100) as indicated. Cells were collected at 18, 23 and 27 h after addition of serum-containing media and processed for flow

cytometry as described²⁶. The horizontal axis reflects relative DNA content and the vertical axis represents cell number. **b**, Either E2F2 expression or cyclin E/Cdk2 restores activation of E2F target genes. Cells treated as in **a** were collected 18 h after serum addition and processed for northern analysis using specific probes for B-Myb, cyclin A, RR2 and GAPDH.

Myc-mediated activation of *E2F1* gene expression involves the direct binding of Myc to the *E2F1* gene promoter, it is clear from this experiment that Myc can activate *E2F* gene expression, leading to the accumulation of E2F activity independent of Rb phosphorylation and cell-cycle progression. This result is also consistent with recent findings that demonstrate a requirement for cyclin E-dependent kinase activity in the induction of S phase that is independent of its role in Rb phosphorylation and E2F activation^{9,13,14}.

The mechanism by which Myc and Ras collaborate to induce cyclin E-dependent kinase activity could involve the activation of kinase activity directly or the inactivation of inhibitors of kinase activity. As shown in Fig. 4, at least part of the effect is probably due to a loss of inhibition, because expression of Myc and Ras in quiescent REF52 cells led to a loss of the p27 CKI protein, whereas

expression of either Myc or Ras alone had no effect on p27 levels. This result is consistent with previous work showing a loss of p27 in response to interleukin-2 or expression of Myc¹⁵⁻¹⁷. The decrease in p27 protein levels in cells expressing Myc and Ras was again not an indirect consequence of cell-cycle progression as coexpression of p21, which inhibited both cyclin E-dependent kinase activity and S-phase entry, had no effect on the Myc/Ras-induced decline in p27. The loss of p27 protein seems to be a post-transcriptional event because p27 mRNA levels did not vary as a function of Myc and Ras expression (Fig. 4). We thus conclude that the activation of G1 CDK activity as a function of Myc and Ras expression may be due in part to the loss of inhibition of kinase activity.

These experiments describe a complex and synergistic relationship between signal transduction activities known to be important

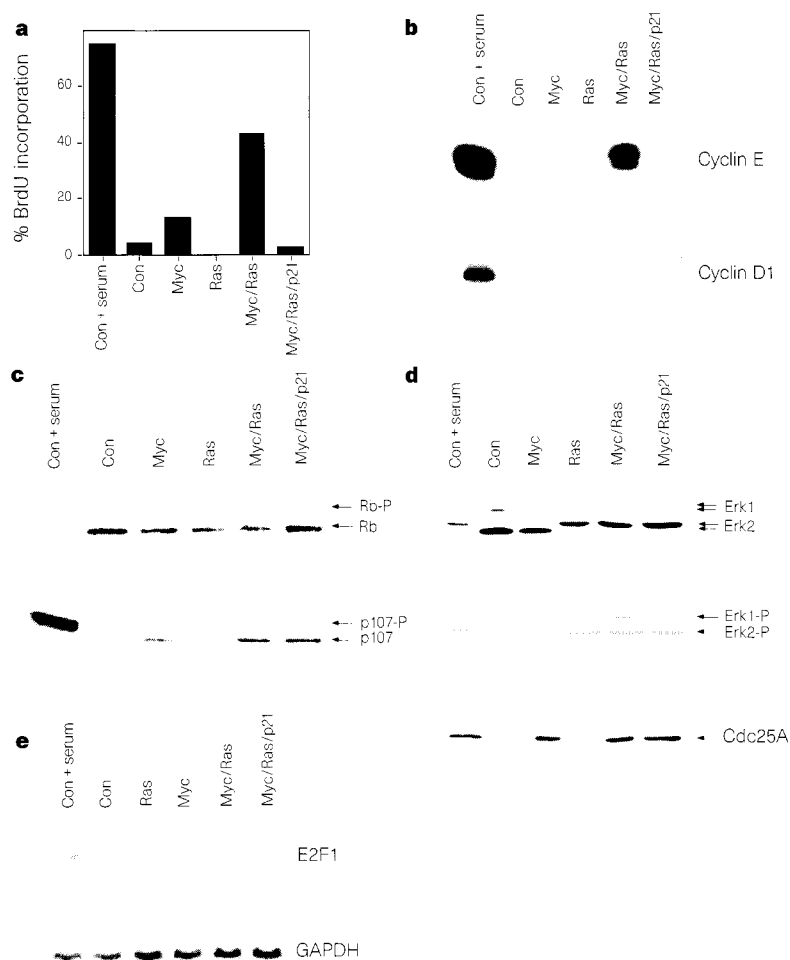


Figure 3 Ras and Myc collaborate to allow accumulation of G1 CDK activity and induction of S phase. **a**, Quiescent REF52 cells were infected with the indicated viruses and BrdU was added after 12 h of infection. Cells were collected 30 h post-infection and BrdU detected by indirect immunofluorescence. Multiplicities of infection were: Ad-Myc (200); Ad-Ras61L (100); Ad-p21 (200); and Ad-Con (300). **b**, Expression of Ras and Myc leads to induction of cyclin E/Cdk2 but not cyclin D/Cdk4 activity. Cells treated as in **a** were collected 30 h post-infection, lysates were immunoprecipitated, and associated kinase activity was determined using histone H1 (top) and GST-Rb (bottom) as previously described²⁷. Affinity-purified rabbit polyclonal antibodies anti-cyclin E (SC-481) and anti-cyclin D1 (SC-450) were purchased from Santa Cruz Biotechnology. **c**, Expression of Ras and Myc does not result in Rb or p107 phosphorylation. Cells treated as in **a** were collected 30 h post-infection in SDS loading buffer and Rb (top) and p107 (bottom) proteins were detected by western blotting using anti-Rb monoclonal antibody 14001A

(Pharmingen) and anti-p107 rabbit polyclonal antibody SC-318 (Santa Cruz Biotechnology) as described²⁴. **d**, Coexpression of Ras and Myc does not affect the activity of either protein. Cells were collected 30 h post-infection in SDS loading buffer and Erk1 and Erk2 were detected by western blotting using an anti-Erk1 monoclonal antibody (which cross-reacts with Erk2) M12320 (top) (Transduction Laboratories) or with a phospho-specific rabbit polyclonal anti-Erk1/2 antibody 9101A (middle) (New England Biolabs). The phosphorylated forms of Erk1 and Erk2 represent the upper band of each doublet; the Erk1 protein in the Con + serum sample is present but it is underrepresented in this sample. Cdc25A protein was detected by western blotting using rabbit polyclonal anti-cdc25A antibody SC-97 (bottom) (Santa Cruz Biotechnology). **e**, Myc activates the expression of the *E2F1* gene. Cells treated as in **a** were collected after 15 h infection and processed for northern analysis using specific probes for E2F1 and GAPDH.

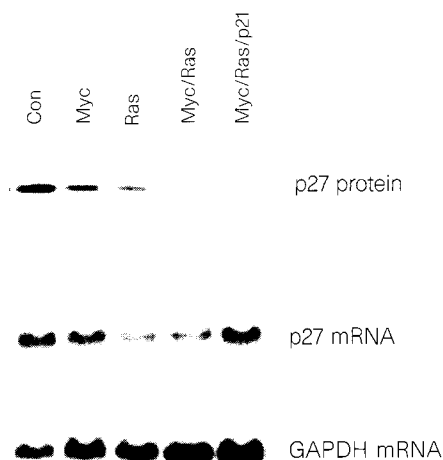


Figure 4 Expression of Myc and Ras leads to a loss of the p27 cyclin-dependent kinase inhibitor. REF52 cells were plated, starved and infected with the indicated viruses as in Fig. 3. Cells were either collected in SDS loading buffer at 30 h post-infection and p27 protein was detected by western blotting using rabbit polyclonal anti-p27 antibody SC-528 (Santa Cruz) (top row) or collected and processed for northern analysis using specific probes for p27 and GAPDH mRNAs (middle and bottom rows).

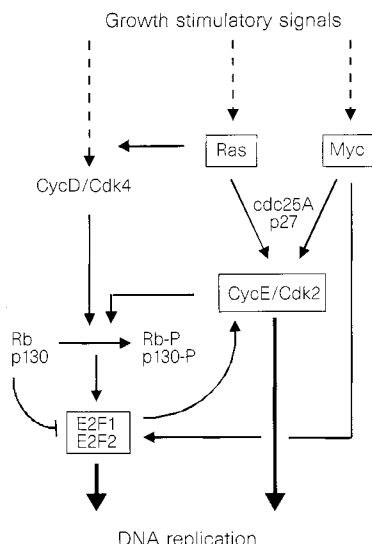


Figure 5 A cooperative role for Ras and Myc in the G1 regulatory pathway. The diagram depicts multiple mechanisms by which Ras, together with Myc, can contribute to the activation of cyclin E/Cdk2 and E2F activities, two activities essential for the induction of S phase. From these experiments and others, Ras is seen to be required for the induction of cyclin D- and cyclin E-dependent kinase activity. This leads in part to the phosphorylation of Rb and Rb-related proteins allowing the accumulation of E2F activity. Myc is also seen to facilitate the activation of the *E2F1* gene, as shown here, as well as the *E2F2* gene (R. Sears *et al.*, unpublished data). In addition, Ras and Myc are seen to collaborate in the induction of cyclin E-dependent kinase activity. In part, this seems to involve the ability of Myc to activate the *cdc25A* gene as well as the ability of Raf, a downstream effector of Ras activity, to activate *cdc25A* function. Moreover, as we show here, Ras and Myc collaborate to eliminate the p27 CKI.

for cell growth. In particular, we show that Myc cooperates with the Ras pathway to generate cyclin E-dependent kinase activity and to allow S-phase induction. This result, together with several additional observations, suggests multiple mechanisms by which Ras, together with Myc, may contribute to the activation of cyclin E/Cdk2 activity (Fig. 5). First, recent experiments have shown that the *cdc25A* gene is a target for transcriptional activation by Myc¹². Thus, the collaboration we observe involving Ras and Myc could represent the Myc-dependent increase in Cdc25A protein¹² coupled with a Raf-dependent activation of Cdc25A phosphatase activity¹⁸. Second, our results demonstrate a dramatic decrease in the levels of the p27 CKI protein in response to the combined action of Myc and Ras. Given the ability of p27 to block cyclin E/Cdk2 activity, the role of Ras and Myc in facilitating the accumulation of kinase activity could be seen as a synergistic process involving both the activation of the kinase through Cdc25A action and the inhibition of p27-mediated negative control of kinase activity. It is also clear from these experiments that a complex relationship exists between Myc and Ras and the activation of E2F. In part, this involves the stimulation of G1 CDK activity leading to phosphorylation of Rb and Rb-related proteins, thereby eliminating the E2F-Rb complex that represses transcription of the *E2F1* and *E2F2* genes in quiescent cells. But the connection goes further, as Myc has the ability to activate *E2F1* gene expression. In addition, analysis of the *E2F2* promoter has revealed the presence of multiple Myc-binding sites that are essential for activation by Myc and for full promoter activity following stimulation of cell growth (R. Sears *et al.*, unpublished data). Finally, the experiments presented here, demonstrating a cooperative relationship between the action of Ras and Myc in the G1 CDK/Rb/E2F pathway, contribute to a molecular interpretation of the many previous experiments that have shown a cooperation between these two genes in the oncogenic transformation of primary fibroblasts and the REF52 cell line^{7,8}. □

Methods

Cells and viruses. Viral stocks were created and virus purified as described previously¹⁹. Viral titres were determined by an indirect immunofluorescent assay specific for the viral 72K E2 gene product as described²⁰ and defined as focus forming units (FFU) per ml. REF52 cells were grown in D-MEM containing 5% fetal bovine serum (FBS) and 5% calf serum. To bring cells to quiescence, REF52 cells were plated at approximately 3,500 cells cm⁻² and incubated overnight. The next day, the cells were washed once with D-MEM and the culture medium replaced with D-MEM containing 0.25% serum and cells were further incubated for 48 h before virus infection. Cells on plates were infected in D-MEM with 20 mM HEPES, pH 7.2, for 75 min at 37 °C at a cell-to-volume ratio of 0.5 × 10⁶ cells ml⁻¹. After infection, four volumes of 0.25% serum/D-MEM were added to each plate, and the cells were incubated at 37 °C. Where indicated, the cells were subsequently serum-stimulated by replacement with culture media containing 10% serum.

The construction of Ad-E2F1, Ad-Con, Ad-CycE, Ad-Cdk2, Ad-E2F2, and Ad-p21 has been described previously (R. Sears *et al.*, unpublished data; refs 9, 21). Ad-Ras^{N17} and Ad-Ras^{61L} were similarly constructed by ligation of the *HindIII/XbaI* fragment from pRsaα Ras^{N17} (ref. 22) and the *BamHI* fragment from pZIP-NeoSV(x)²³ into the similarly digested pGEM-CMV vector. Ad-Myc was constructed by ligation of the *HindIII/XbaI* fragment from CMV-Myc containing the mouse *c-myc* cDNA into the similarly digested pGEM-CMV vector.

Received 19 December 1996; accepted 4 April 1997.

- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G₁ cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
- Eisenman, R. N. & Cooper, J. A. Signal transduction: Beating a path to Myc. *Nature* **378**, 438–439 (1995).
- Mulcahy, L. S., Smith, M. R. & Stacey, D. W. Requirement for *ras* proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* **313**, 241–243 (1985).
- Lowry, D. R. & Willumsen, B. M. Function and regulation of *ras*. *Annu. Rev. Biochem.* **62**, 851–891 (1993).
- Feig, L. A. & Cooper, G. M. Inhibition of NIH 3T3 cell proliferation by a mutant *ras* protein with preferential affinity for GDP. *Mol. Cell. Biol.* **8**, 3235–3243 (1988).

7. Land, H., Parada, L. F. & Weinberg, R. A. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* **304**, 596–602 (1982).
8. Kohl, N. E. & Ruley, H. E. Role of c-myc in the transformation of REF52 cells by viral and cellular oncogenes. *Oncogene* **2**, 41–48 (1987).
9. Leone, G., DeGregori, J., Jakoi, L. & Nevins, J. R. Collaboration of E2F and cyclin E/cdk2 in the induction of S phase. *Proc. Natl Acad. Sci. USA* (submitted).
10. Hatekeyama, M., Brill, J. A., Fink, G. R. & Weinberg, R. A. Collaboration of G₁ cyclins in the functional inactivation of the retinoblastoma protein. *Genes Dev.* **8**, 1759–1771 (1994).
11. Resnitzky, D. & Reed, S. I. Different roles for cyclins D1 and E in regulation of the G₁-to-S transition. *Mol. Cell. Biol.* **15**, 3463–3469 (1995).
12. Galaktionov, K., Chen, X. & Beach, D. Cdc25 cell-cycle phosphatase as a target of c-myc. *Nature* **382**, 511–517 (1996).
13. Duronio, R. J., O'Farrell, P. H., Xie, J.-E., Brook, A. & Dyson, N. The transcription factor E2F is required for S phase during *Drosophila* embryogenesis. *Genes Dev.* **9**, 1445–1455 (1995).
14. Duronio, R. J., Brook, A., Dyson, N. & O'Farrell, P. H. E2F-induced S phase requires cyclin E. *Genes Dev.* **10**, 2505–2513 (1996).
15. Steiner, P. et al. Identification of a Myc-dependent step during the formation of active G₁ cyclin-cdk complexes. *EMBO J.* **14**, 4814–4826 (1995).
16. Coats, S., Flanagan, W. M., Nourse, J. & Roberts, J. M. Requirement of p27^{Kip1} for restriction point control of the fibroblast cell cycle. *Science* **272**, 877–880 (1996).
17. Nourse, J. et al. Interleukin-2-mediated elimination of the p27^{Kip1} cyclin-dependent kinase inhibitor prevented by rapamycin. *Nature* **372**, 570–573 (1994).
18. Galaktionov, K., Jessus, C. & Beach, D. Raf1 interaction with Cdc25 phosphatase ties mitogenic signal transduction to cell cycle activation. *Genes Dev.* **9**, 1046–1058 (1995).
19. Nevins, J. R., DeGregori, J., Jakoi, L. & Leone, G. Functional analysis of E2F. *Methods Enzymol.* (in the press).
20. DeGregori, J., Kowalik, T. & Nevins, J. R. Cellular targets for activation by the E2F1 transcription factor include DNA synthesis and G1/S regulatory cells. *Mol. Cell. Biol.* **15**, 4215–4224 (1995).
21. Schwarz, J. K. et al. Expression of the E2F1 transcription factor overcomes type β transforming growth factor-mediated growth suppression. *Proc. Natl Acad. Sci. USA* **92**, 483–487 (1995).
22. Farnsworth, C. L. & Reig, L. A. Dominant inhibitory mutations in the Mg²⁺-binding site of Ras^H prevent its activation by GTP. *Mol. Cell. Biol.* **11**, 4822–4829 (1991).
23. Khosravi-Far, R., Solis, P. A., Clark, G. J., Kinch, M. S. & Der, C. J. Activation of Rac1, RhoA, and mitogen-activated protein kinases is required for Ras transformation. *Mol. Cell. Biol.* **15**, 6443–6453 (1995).
24. Ikeda, M.-A., Jakoi, L. & Nevins, J. R. A unique role for the Rb protein in controlling E2F accumulation during cell growth and differentiation. *Proc. Natl Acad. Sci. USA* **93**, 3215–3220 (1996).
25. Jacks, T. et al. Effects of an Rb mutation in the mouse. *Nature* **359**, 295–300 (1992).
26. Smith, E. J., Leone, G., DeGregori, J., Jakoi, L. & Nevins, J. R. The accumulation of an E2F-p130 transcriptional repressor distinguishes a G₀ from a G₁ cell state. *Mol. Cell. Biol.* **16**, 6965–6976 (1996).
27. DeGregori, J., Leone, G., Ohtani, K., Miron, A. & Nevins, J. R. E2F1 accumulation bypasses a G₁ arrest resulting from the inhibition of G₁ cyclin-dependent kinase activity. *Genes Dev.* **9**, 2873–2887 (1995).

Acknowledgements. We thank M. Cook and A. Fisher of the Duke Comprehensive Cancer Center Flow Cytometry Facility for the flow cytometric analysis; T. Jacks for the Rb^{+/+} and Rb^{-/-} fibroblasts; C. Der for the H-Ras^{WT} cDNA; B. Lefkowitz for the H-Ras^{WT} cDNA; K. Culler for assistance in preparing the manuscript; and R. Rempel for critical comments. G.L. is supported by fellowships from the Medical Research Council of Canada (MRC) and the Alberta Heritage Foundation for Medical Research; J.D. is supported by the HHMI; and R.S. by an NIH postdoctoral fellowship.

Correspondence and requests for materials should be addressed to J.R.N. (e-mail: nevin001@mc.duke.edu).

Atomic structure of the ectodomain from HIV-1 gp41

W. Weissenhorn^{*†‡}, A. Dessen^{*†‡}, S. C. Harrison^{*†‡}, J. J. Skehel[§] & D. C. Wiley^{*†‡}

^{*} Laboratory of Molecular Medicine and [†] Howard Hughes Medical Institute, The Children's Hospital, 320 Longwood Avenue, Boston, Massachusetts 02215, USA

[‡] Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

[§] National Institute of Medical Research, Mill Hill, The Ridgeway, London NW7 1AA, UK

|| These authors contributed equally to this work

Fusion of viral and cellular membranes by the envelope glycoprotein gp120/gp41 effects entry of HIV-1 into the cell. The precursor, gp160, is cleaved post-translationally into gp120 and gp41 (refs 1, 2), which remain non-covalently associated. Binding to both CD4 and a co-receptor leads to the conformational changes in gp120/gp41 needed for membrane fusion³. We used X-ray crystallography to determine the structure of the protease-resistant part^{4,5} of a gp41 ectodomain solubilized with a trimeric GCN4 coiled coil in place of the amino-terminal fusion peptide⁶. The core of the molecule is found to be an extended, triple-stranded α -helical coiled coil with the amino terminus at its tip. A

carboxy-terminal α -helix packs in the reverse direction against the outside of the coiled coil, placing the amino and carboxy termini near each other at one end of the long rod. These features, and the existence of a similar reversal of chain direction in the fusion pH-induced conformation of influenza virus HA2 (ref. 7) and in the transmembrane subunit of Moloney murine leukaemia virus⁸ (Fig. 1a–d), suggest a common mechanism for initiating fusion.

A soluble crystalline GCN4/gp41 chimera (Fig. 1) composed of the first 31 residues of a trimeric GCN4 isoleucine zipper⁹, the proposed zipper domain of gp41, residues 30–79 (refs 10, 11), and the C-terminal helical segment of gp41, residues 113–154 (refs 4, 5, 11), was produced by expressing the two segments in bacteria, refolding *in vitro*, and trimming with protease⁶. It lacks the 33 proteolytically sensitive residues 80–112, which contain a short disulphide loop and two oligosaccharide sites. It also lacks the 18 C-terminal residues of the gp41 ectodomain^{4–6}. Electron microscopy indicates that the recombinant chimera lacking these protease-sensitive regions has the same elongated, rod-like shape as an intact gp41 ectodomain expressed and secreted in insect cells¹². The structure was determined by isomorphous replacement. The final model contains residues 1–77 and 117–154 and 15 solvent molecules.

The structure is a strikingly regular trimeric rod (Fig. 2a, b). There is a central α -helical coiled coil (residues 1–77) 115 Å long and an antiparallel outer α -helical layer (residues 117–154) 60 Å long. The α -helices are packed together in the classical 'knobs into holes' arrangement¹³, but the outer layer packs differently onto the core from the way predicted by helical wheel sequence analysis and biochemical data^{4,5}. Along the length of the central three-stranded coiled coil there are 15 layers of homotrimeric interactions in which successive 'a' and 'd' positions of the heptad repeats (A30L34IQ-LIQLTILIVLQ79; Fig. 1e) pack together on the molecular threefold symmetry axis. The C-terminal helices (117–154) run antiparallel from opposite position 65 of the core helices back to position 30 (Fig. 2b). Each C-terminal helix follows approximately the groove between two core helices. The connection of a core helix to a C-terminal helix is not present in our structure. The shortest path from the C terminus of one to the N terminus of the other links a core helix to its anticlockwise C-terminal neighbour, viewed from the 'fusion peptide' (Fig. 2a, b). This direction for the linker agrees with the direction of chain reversal seen in the influenza HA2 fusion pH structure and in the transmembrane (TM) subunit of Moloney murine leukaemia virus (MoMuLV) (Fig. 1a–c).

Because they run along a groove but lie at a larger radius, the outer-layer helices are tilted with respect to the core helices by approximately 15 degrees. Typical layers of packed side chains are shown in Fig. 2c, d. The relative tilt of core and outer-layer helices leads to a gradual drift in axial registration of the layers so that there is some irregularity near the 'bottom' of the bundle (Fig. 2 legend). A cluster of conserved tryptophan residues (W60, W117 and W120) dominate the helical packing in this region, and their bulk increases the interhelix spacing (Fig. 2c). Near the midpoint of the structure there is a cluster of conserved glutamine residues, and a network of hydrogen bonds links several glutamine and asparagine side chains near residue 40 in the core with a similar cluster near residue 142 in the outer layer (Fig. 2d). The Gln 41 side chains, at 'a' positions, face each other across the threefold axis. Several other hydrogen bonds and salt bridges seem to stabilize the observed structure. Most of these non-polar and polar interactions are from conserved or conservatively substituted residues in HIV-1, HIV-2 and simian immunodeficiency virus (SIV) (Fig. 1e).

The coiled-coil core extends beyond the outer helical layer at both ends (Fig. 2a, b). At its C-terminal end (residue 77, bottom in Fig. 2a), the overhang is five helical turns, or about 25 Å. This overhang would be linked to an outer-layer helix by 39 residues. The missing linker would contain a disulphide loop and two conserved glycosylation

sites. The outer-layer helices begin at residue 117. In complete gp41 the helices probably begin at approximately this point because there is a glycosylation site at the preceding heptad position (113) in the HX2B strain.

GCN4/gp41 from HIV-1, the fusion pH-induced conformation of influenza virus HA2 (ref. 7), and the TM fragment of the MoMuLV⁸ share three features of overall molecular organization (Fig. 1a–c). First, the N-terminal fusion peptide is at the tip of a triple-stranded, α -helical coiled coil. In the case of influenza haemagglutinin, it is known that the fusion pH-induced conformational change uncovers the fusion peptide and causes it to move more than 100 Å from the centre of the molecule to one end, as the result of extension of the coiled-coil core by 36 residues^{7,14,15} (Fig. 1c, d). Second, the polypeptide chain reverses direction at the other end of the coiled coil and proceeds back towards the N-terminal tip. The turn occurs at the same place in HIV-1 gp41 and MoMuLV subunit, residue 78, and is followed by a glycine-rich stretch of eight residues and a disulphide loop of seven (HIV-1) or eight (MoMuLV)

residues (Fig. 1b), suggesting further structural similarities between gp41 and the transmembrane components of other retroviruses. Third, the C-terminal link leading to the transmembrane anchor is probably flexible. It is exposed to proteolysis in the retroviral and influenza virus proteins^{4–6,8,16}, and it is structurally disordered in the fusion pH-induced conformation of HA2 (ref. 7).

The flexibly linked transmembrane anchors and the fusion peptides may be constrained to the same end of the coiled coil by the shared features just outlined. The C terminus of the proteolytically resistant core of GCN4/gp41, residue 154, is only 18 residues from the transmembrane anchor, yet it lies nearly at the N-terminal end of the helical rod, over 75 Å from the 'bottom' of the coiled coil (Fig. 2a). In the influenza virus HA2 structure, the last observed residue is 23 residues from the transmembrane anchor, and it also lies close to the N-terminal end of the helical rod, approximately 70 Å from the 'bottom' of the coiled coil (Fig. 1c). The last residue in the crystalline fragment (the central 40%) of the MoMuLV TM subunit (46–97) is 37 residues from the transmembrane anchor and

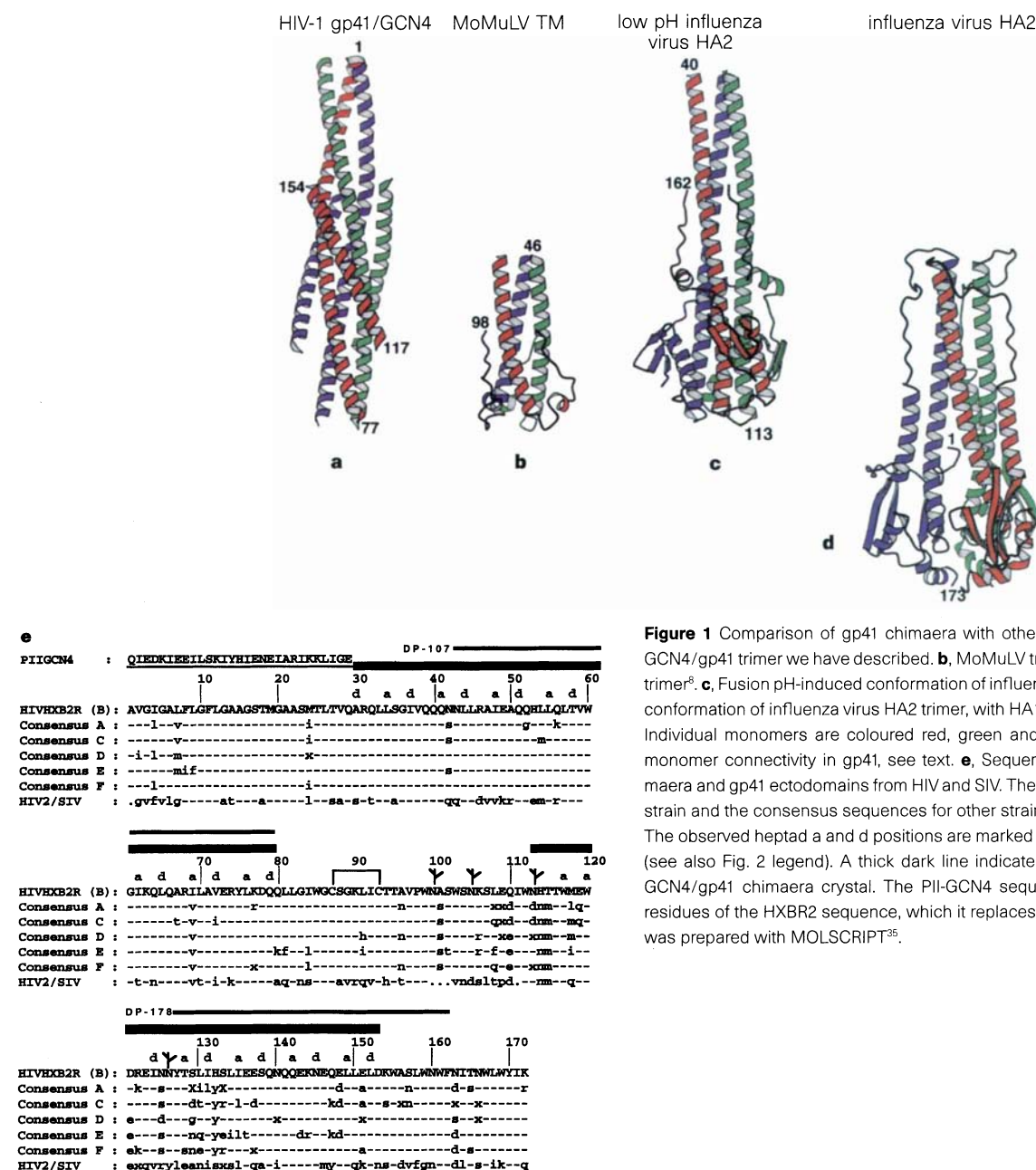


Figure 1 Comparison of gp41 chimaera with other viral glycoproteins. **a**, The GCN4/gp41 trimer we have described. **b**, MoMuLV transmembrane (TM) subunit trimer⁸. **c**, Fusion pH-induced conformation of influenza virus HA2 (ref. 7). **d**, Initial conformation of influenza virus HA2 trimer, with HA1 domains omitted for clarity. Individual monomers are coloured red, green and purple. For assignment of monomer connectivity in gp41, see text. **e**, Sequences of the GCN4/gp41 chimaera and gp41 ectodomains from HIV and SIV. The sequence of the HIV HXB2R strain and the consensus sequences for other strain classes have been listed³⁴. The observed heptad a and d positions are marked above the HXB2R sequence (see also Fig. 2 legend). A thick dark line indicates the portion of gp41 in the GCN4/gp41 chimaera crystal. The PII-GCN4 sequence⁹ is above the first 29 residues of the HXB2R sequence, which it replaces in the chimaera. This figure was prepared with MOLSCRIPT³⁵.

lies roughly midway along the coiled coil formed by residues 46–78 (Fig. 1b). Circular dichroism indicates that part of the C-terminal segment omitted from the crystallized fragment of the MoMuLV TM subunit is α -helical⁸. Taken together these observations suggest that once the outer layer around the central coiled coil has zipped into place, the transmembrane anchors and fusion peptides will be at the same end of the rod-shaped molecules. Electron microscopy and antibody labelling of membrane-associated HA2 provide direct evidence for this suggestion¹⁷.

Binding of the HIV envelope glycoprotein to its receptors results in increased exposure of gp41 and ultimately to the dissociation of gp120 (refs 18–20). Thus we expect gp41 produced in the absence of gp120 to adopt the conformation of its fusion-active state, just as influenza virus HA2 produced in bacteria in the absence of HA1 adopts its fusion pH-induced conformation^{7,21}. The long coiled-coil

structures in recombinant gp41, the MoMuLV TM subunit and HA2 expressed in the absence of their receptor-binding subunits^{5,8,12,21}, and in viral HA2 in the fusion pH-induced conformation¹⁶, are all extremely thermostable. These observations are consistent with the expectation that, once triggered by interaction with the cell or transfer into endosomes, some viral fusion proteins switch irreversibly during fusion from a metastable to a stable conformation.

Evidence that infectivity requires a conformational change in gp41, rather than merely 'uncapping' by dissociation of gp120, comes from the properties of site-directed point mutations in the coiled coil and from the mechanism of inhibition of HIV-1 infectivity by peptides derived from gp41 (refs 22–24). Mutations at Ile 62, an 'a' heptad position in the coiled coil (Figs 1e, 2c), inhibit both membrane fusion and viral infectivity without preventing formation

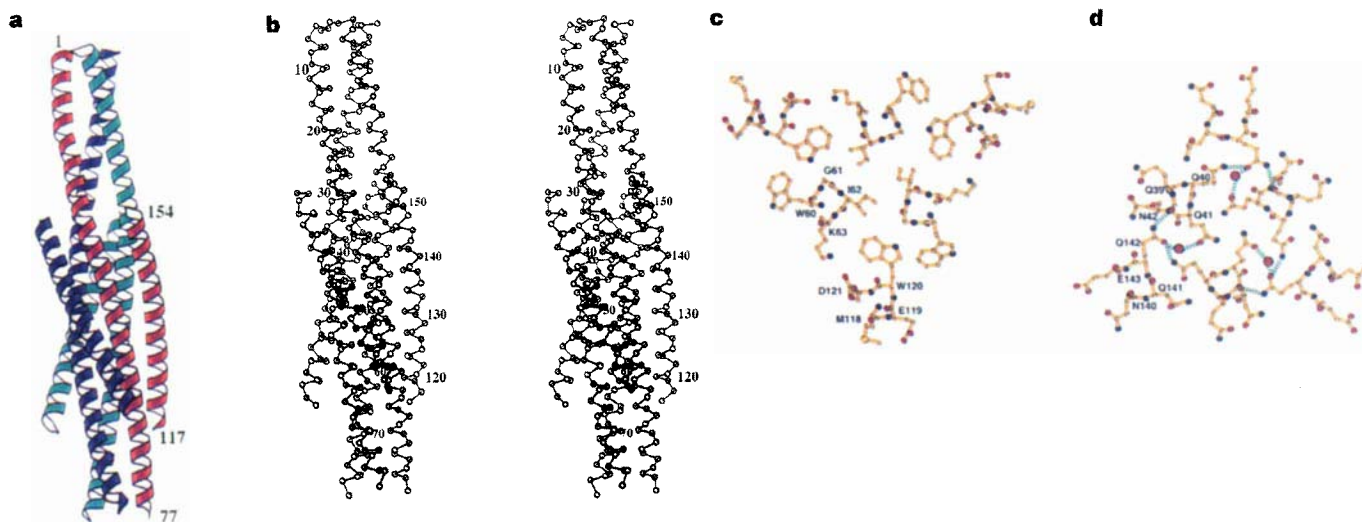


Figure 2 GCN4/gp41 chimera structure. **a**, Ribbon drawing of trimeric GCN4/gp41. **b**, C α stereo drawing of trimeric GCN4/gp41, labelled every 10 residues. **c**, Helix packing in gp41, seen in cross-section at two levels in the structure. Homotrimeric packing of I62, intra-chain packing of W120 against I62 of the same molecule, and W120 against an adjacent core helix in the oligomer. (Labels are all on one monomer.) **d**, Packing in the glutamine (Q)-rich layer around residues 41 and 142 in the core and outer layers, respectively. Hydrogen bonds are shown in green as broken lines. A bound solvent molecule is shown as an isolated red circle, linked by hydrogen bonds to Q141 and Q142. Q141 in the outer layer (and

Y127, L134, E148) makes pseudo-twofold contacts with one of the core helices, not pseudo-threefold contacts as predicted^{4,5}. Successive a and d positions of each outer layer helix (W117 W120 I124 T128 I131 I135 SQNLL152; Figs 1e, 2b) form 11 layers of pseudo-threefold contacts with the two adjacent core helices. Alternate aag and dde heptad layers form this interface except at the first layer (W117), which is dae; there is also one extra residue between 'a' 120 and 'd' 124. (The designations aag and dde refer to a core helix, a C-terminal helix, and a second core helix, respectively, in anticlockwise order as viewed from the GCN4 end.) This figure was prepared using MOLSCRIPT³⁵.

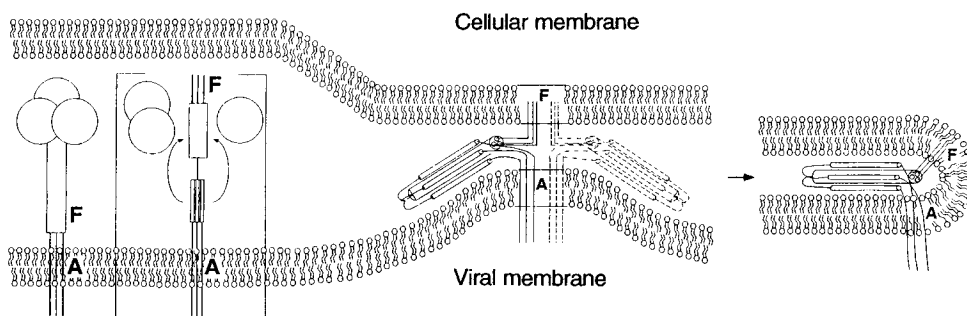


Figure 3 A model for the interaction with membranes of gp41 and other viral fusion subunits, during membrane fusion. Left, before fusion, viral glycoproteins project their receptor-binding domains (spheres) towards the cellular membrane. Brackets, a conformational change extends the N-terminal fusion peptide (F) towards the cellular membrane. Centre, after the outer layer of the fusion conformation has assembled, the N-terminal fusion peptides (F) and the C-terminal transmembrane anchors (A) lie near each other at a site of close

apposition of the prefusion membranes. Flexible links between the central rod and the F and A segments allow variable orientations of the rod with respect to the two membranes. A second trimer, shown in dotted lines, indicates how such trimers might aggregate at their hydrophobic ends at initial sites of fusion. Right, after membrane fusion, both the fusion peptides (F) and transmembrane anchors (A) are shown in the same membrane, as suggested previously^{7,17,26}.

Table 1 Statistics for data collection, phase determination and refinement

Crystal	Native	OsO ₄	K ₂ PtBr ₄	HgI ₄
Resolution limit (Å)	2.6	2.6	3.0	3.3
R _{sym} (%) [*]	7.1	9.7	9.7	9.8
Completeness	90.5	87.9	86.1	91.5
R _{deri} (%) [†]		14.0	15.7	27.0
Sites		2	1	1
Phasing power (acentric/centric) [‡]		1.58/1.24	1.16/0.78	1.22/1.14
R _{cullis} (% acentric/centric) [§]		73/59	84/73	82/74
Refinement				
Resolution	25–2.6 Å			
R _{free} (%)	27.7			
R _{cryst} (%)	24.4			

	Res. No.	Average B-factor	Bonds (Å)	Angles (°)	B-factors (Å ² bonded)
Protein	115	53.1	0.010	1.20	4.5
Water	15	54.2			

^{*}R_{sym} = $\sum ||I - \langle I \rangle|| / \sum I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[†]R_{deri} = $\sum ||F_{p4} - |F_p|| / \sum |F_p|$, where $|F_p|$ is the protein structure factor amplitude and $|F_{p4}|$ is the heavy-atom derivative structure factor amplitude.

[‡]Phasing power = r.m.s. ($|F_H|/|E|$), where $|F_H|$ is the heavy-atom structure factor amplitude, and E is the residual lack of closure error.

[§]R_{cullis} = $\sum ||E| - |F_p|| / \sum |F_p|$.

^{||}R = $\sum ||F_o| - |F_c|| / \sum |F_o|$, where R_{free} is calculated for a randomly chosen 10% of reflections, and R_{cryst} is calculated for the remaining 90% of reflections ($F > 2.0$) used for structure refinement.

of gp120/gp41 oligomers²⁵. Ile 62 makes a homotrimeric contact at the centre of the coiled coil and interacts with Trp 120 of the outer helix, in the cluster of conserved tryptophans (Fig. 2c). Polar mutations at 62 would destabilize both the coiled coil itself and its association with the outer helix, potentially exerting a greater effect on the fusion state of gp41 than on the oligomerization of gp120/gp41 (refs 23, 25), thus inhibiting the conformational transition.

Peptides having the sequence of residues 42–72 and 127–162 from gp41 (Fig. 1e) have anti-HIV activity and block membrane fusion^{22–24}. In GCN4/gp41 these regions interact directly: 42–72 is part of the coiled-coil sequence against which the outer helix segment 127–162 packs. We would not expect these peptides to interact with the structure of gp41 seen here, which is too stable to be disrupted by their binding. Indeed, our structure supports the proposal that these peptides exert their effects by interacting with gp41 during the conformational change to the fusion-active conformation²⁴. Only before or during that conformational change would we expect the targets for the peptides to be available. Peptide 42–72 could target the incipient outer-layer helix and peptide 127–162 the surface of the awaiting coiled coil. This interpretation is consistent with the lack of inhibitory action observed with a preassembled complex containing both the coiled coil and the outer-layer helices⁵. Whatever the actual inhibitory mechanism of these peptides, it should indeed be possible to design or discover small-molecule drugs that reduce the infectivity of HIV, influenza or other viruses by interfering with chain reversal, as suggested here.

By analogy with the changes in influenza haemagglutinin structure induced at fusion pH, the conformational change in gp41 required for activation of membrane fusion by association of the envelope glycoprotein with its receptor might also include formation of the complete coiled coil by extension at the N terminus. For example, the glutamine-rich segments near residues 41 and 51 described earlier (Fig. 1e) might not be α -helical in native gp120/gp41. Like the 'B-loop' in influenza HA2, which undergoes a loop-to-helix transition at the fusion pH (refs 7, 15), gp41 is particularly rich in polar and charged side chains between residues 39 and 53.

The observation that conformational changes in a membrane fusion protein uncover its fusion peptide has led to the expectation that the exposed hydrophobic peptide would initiate fusion by inserting into a membrane, probably the cellular membrane^{14,26}. The

observation that, during its fusion pH-induced conformational change, influenza HA2 folds back and assembles an outer layer around a central coiled coil⁷ has led to the notion that in the postfusion state there would be an inverted molecule with both ends in the same membrane^{7,17}. Flexible linkages between the central rod-like segment and the membrane insertions at either end would allow the molecule to bend as the outer layer forms, thereby bringing the apposed membranes more closely together than the length of the rod itself^{17,26}. The structure of GCN4/gp41 suggests, more strikingly than previous structures, that the fusion peptides and transmembrane anchors of the fusion glycoproteins are likely to be juxtaposed at one end of the coiled coil, once the outer layer of helices has assembled onto the core. This arrangement would promote a close apposition of viral and cellular membranes, as suggested in Fig. 3. The transmembrane anchors and fusion segments on the envelope protein of tick-borne encephalitis and other flaviviruses may also be juxtaposed near one end of the extended molecule, which even in the native virion lies parallel to the viral membrane²⁷.

The presence of the hydrophobic fusion peptide and transmembrane anchors in two bilayers in direct apposition may be central to overcoming the 'hydration force'²⁸ that is a barrier to membrane fusion. The nonlinear dependence on fusion-protein surface density of a kinetic lag in the reaction suggest that several trimers may assemble to create this pore²⁹. It is possible that long, rod-shaped molecules may cluster at their hydrophobic tips (Fig. 3), forming asters with centres that could be the points of initial membrane fusion (see, for example, ref. 30). Since this paper was submitted, the structure of a smaller fragment of gp41 has been reported³¹. □

Methods

Diffraction data. The GCN4/gp41 chimera pII41NC(113–154) was prepared as described⁶. Crystals were grown in hanging drops by combining 1 μ l protein solution (15–20 mg ml⁻¹ in 20 mM HEPES, pH 8.3, 75 mM NaCl) with 1 μ l of reservoir solution (100 mM HEPES, pH 8.0, 1.1 M ammonium sulphate, 12% ethylene glycol). They belong to space group R32, $a = 52.37$ Å, $b = 52.37$ Å, $c = 414.52$ Å, one monomer per asymmetric unit. Diffraction data were recorded at room temperature, using a Mar image plate detector and an Elliot GX-13 rotating anode source with mirror optics, and processed using the programs DENZO and SCALEPACK (HKL Research).

Structure determination. The structure was determined by multiple isomorphous replacement (MIR). For derivatization, crystals were soaked in 0.5 mM K₂OsO₄ for 48 h, in 200 mM K₂PtBr₄ for 7 days, and in 1 mM K₂HgI₄

for 60 h at room temperature. Native and derivative data were scaled using CCP4 programs³². One Os site was located in a difference Patterson map; an additional Os site, one Pt site, and one Hg site were found in difference Fourier maps. Heavy-atom positions were refined and phases were calculated using MLPHARE³². Both an SIR map calculated solely with Os phase information (including anomalous scattering) and an MIR map, calculated with all derivatives, were improved by density modification using the program DM³². These maps allowed assignment of 92 of the 121 residues of the GCN4/gp41 molecule in the density, using the program O (DATAONO AB). Combined model and MIR phases led to an improved MIR map, which then revealed all except 6 of the remaining residues. Cycles of rebuilding, positional refinement and thermal-parameter refinement³³ were used to improve the model, as judged by the free R -factor³³. The model was submitted to simulated annealing from 3,000K after R_{free} had dropped below 34%, after which further rebuilding and refinement yielded a model with $R_{\text{free}} = 27.7\%$ and $R_{\text{cryst}} = 24.4\%$ to 2.6 Å (all data): see Table 1. The refined model includes residues 1–77 in the long helix and 117–154 in the short helix, as well as 15 water molecules. The four N-terminal amino acids of the short helix, 113–116, did not have identifiable density in difference maps; they may be poorly ordered. Despite high thermal parameters in the GCN4 region, the chain is clear throughout the entire model.

Received 16 April; accepted 1 May 1997.

- Allan, J. S. *et al.* Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science* **228**, 1091–1094 (1985).
- Veronese, F. D. *et al.* Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science* **229**, 1402–1405 (1985).
- D'Souza, M. P. & Harden, V. A. Chemokines and HIV-1 second receptors: confluence of two fields generates optimism in AIDS research. *Nat. Med.* **2**, 1293–1300 (1996).
- Blacklow, S. C., Lu, M. & Kim, P. S. A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry* **34**, 14955–14962 (1995).
- Lu, M., Blacklow, S. C. & Kim, P. S. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct. Biol.* **2**, 1075–1082 (1995).
- Weissenhorn, W. *et al.* Assembly of a rod-shaped chimera of a trimeric GCN4 zipper and the HIV-1 Gp41 ectodomain expressed in *E. coli*. *Proc. Natl Acad. Sci. USA* (in the press).
- Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
- Fass, D., Harrison, S. C. & Kim, P. S. Retrovirus envelope domain at 1.7 Å resolution. *Nature Struct. Biol.* **3**, 465–469 (1996).
- Harbury, P. B., Kim, P. S. & Alber, T. Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83 (1994).
- Chambers, P., Pringle, C. R. & Easton, A. J. Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J. Gen. Virol.* **71**, 3075–3080 (1990).
- Gallagher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C. & Montelaro, R. C. A general model for the transmembrane proteins of HIV and other retroviruses. *AIDS Res. Hum. Retroviruses* **5**, 431–440 (1989).
- Weissenhorn, W. *et al.* The ectodomain of HIV-1 env subunit gp41 forms a soluble, alpha-helical, rod-like oligomer in the absence of gp120 and the N-terminal fusion peptide. *EMBO J.* **15**, 1507–1514 (1996).
- Crick, F. H. C. The Fourier transform of a coiled-coil. *Acta Crystallogr.* **6**, 685 (1953).
- Skehel, J. J. *et al.* Changes in the conformation of influenza virus haemagglutinin at the pH optimum of virus-mediated membrane fusion. *Proc. Natl Acad. Sci. USA* **79**, 968–972 (1982).
- Carr, C. M. & Kim, P. S. A spring-loaded mechanism for the conformational change of influenza haemagglutinin. *Cell* **73**, 823–832 (1993).
- Ruigrok, R. W. *et al.* Studies on the structure of the influenza virus haemagglutinin at the pH of membrane fusion. *J. Gen. Virol.* **69**, 2785–2795 (1988).
- Wharton, S. A. *et al.* Electron microscopy of antibody complexes of influenza virus haemagglutinin in the fusion pH conformation. *EMBO J.* **14**, 240–246 (1995).
- Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* **250**, 1139–1142 (1990).
- Hart, T. K. *et al.* binding of soluble CD4 proteins to human immunodeficiency virus type1 and infected cells induces release of envelope glycoprotein gp120. *Proc. Natl Acad. Sci. USA* **88**, 2189–2193 (1991).
- Kirsh, R. *et al.* Morphometric analysis of recombinant soluble CD4-mediated release of the envelope glycoprotein gp120 from HIV-1. *AIDS Res. Hum. Retroviruses* **6**, 1209–1212 (1990).
- Chen, J. *et al.* A soluble domain of the membrane-anchoring chain of influenza virus haemagglutinin (HA2) folds in *Escherichia coli* into the low-pH-induced conformation. *Proc. Natl Acad. Sci. USA* **92**, 12205–12209 (1995).
- Wild, C. T., Oas, T., McDanal, C. B., Bolognesi, D. & Matthews, T. A synthetic inhibitor of human immunodeficiency virus replication: Correlation between solution structure and viral inhibition. *Proc. Natl Acad. Sci. USA* **89**, 10537–10541 (1992).
- Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B. & Matthews, T. J. Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 and gp41 are potent inhibitors of virus infection. *Proc. Natl Acad. Sci. USA* **91**, 9770–9774 (1994).
- Chen, C. H., Matthews, T. J., McDanal, C. B., Bolognesi, D. P. & Greenberg, M. L. A molecular clasp in the human immunodeficiency virus (HIV) type 1 TM protein determines the anti-HIV activity of gp41 derivatives: implication for viral fusion. *J. Virol.* **69**, 3771–3777 (1995).
- Dubay, J. W., Roberts, S. J., Brody, B. & Hunter, E. Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. *J. Virol.* **66**, 4748–4756 (1992).
- Skehel, J. J. *et al.* Membrane Fusion by Influenza Hemagglutinin. *Cold Spring Harb. Symp. Quant. Biol.* **60**, 573–580 (1996).
- Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375**, 291–298 (1995).
- Leikin, S., Parsagian, V. A., Rau, D. C. & Rand, R. P. Hydration forces. *Annu. Rev. Phys. Chem.* **44**, 369–395 (1993).
- Danieli, T., Pelletier, S. L., Henis, Y. I. & White, J. M. Membrane fusion mediated by the influenza virus haemagglutinin requires the concerted action of at least three haemagglutinin trimers. *J. Cell Biol.* **133**, 559–569 (1996).
- Kanaseki, T., Kawasaki, K., Murata, M., Ikeuchi, Y. & Ohnishi, S.-I. Structural features of membrane fusion between influenza virus and liposome as revealed by quick-freezing electron microscopy. *J. Cell Biol.* (in the press).
- Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
- CCP4 The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
- Brünger, A. T. *X-PLOR (Version 3.1) Manual* (Yale University, New Haven, CT, 1992).
- Myers, G. *et al.* in *Theoretical Biology and Biophysics Group, Los Alamos* (National Library, Los Alamos, NM, 1995).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 924–950 (1991).

Acknowledgements. We thank M. Pietras for technical assistance, and present and former members of the Harrison and Wiley groups, including M. Lawrence, R. Nolte, B. Temple, K. Smith, A. Musacchio, W. Xu, M. Eck, P. Rosenthal, R. Brown and S. Ray, for advice. The research was supported by grants from the NIH, the Howard Hughes Medical Institute (HHMI), and the MRC. S.C.H. and D.C.W. are investigators in the HHMI, and W.W. is an HHMI associate.

Correspondence and requests for materials should be addressed to D.C.W. (e-mail: wiley@xta10.harvard.edu). Coordinates will be deposited in the Brookhaven Protein Data Bank and are available pre-release at dessen@xtal16.tch.harvard.edu.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Green pastures for plant scientists

Competitive job markets and the shifting sands of funding in the United States and Europe may prompt agricultural and plant researchers to look at industry, academic-industry collaborations or a stint in a developing country.

Brendan Horton

According to a survey of students done some years ago by DowElanco, an Indianapolis-based joint venture between Dow Chemical and Eli Lilly and Co., many students felt that scientists taking jobs in industry were not of the highest calibre and were 'money grabbers'. Their teachers, however, thought that there was a high level of interest in industrial research and that students should consider postdoctoral programmes in industry.

On average, an academic job posting at the University of Maryland's Center for Agricultural Biotechnology receives 400–500 applications. At Rockefeller University, where there is a collection of independent laboratories instead of the traditional department set-up, Nam-Hia Chua says he typically receives around 200 applications per job listing. Bruce Tabashnik, head of entomology at the University of Arizona, expects to receive 50 applications for one functional ecologist's position that is open in his department. Tabashnik feels that fewer than half of today's new PhDs will find permanent academic jobs. Consequently, job seekers may do well to look again at industry as a career choice.

Changes in industry

One need look no further than the company Monsanto (Life Sciences) or the formation of DowElanco in 1989 to see how industry is changing. Both companies have a team-based approach to research and development, which they believe is the most efficient way to turn ideas into products. At Monsanto, this reform started taking shape in 1987, when company officials began to consider the importance of group dynamics in their hiring practices. According to Steve Padgett, programme director for Monsanto's soybean group, the team concept progressed with the further development of research teams, later followed by the development of teams on the business side and then around specific crops.

Given the emphasis on team-based research, Janice Edwards, director for carbohydrate technology, stresses the importance of oral as opposed to written communication, for example in presenting an idea for a research project to Monsanto. This involves an exchange at different levels of the organization, she says, so that when a project review is scheduled the ideas have been consolidated. To be considered for a job in industry, a candidate needs "good academic credentials to get in the door, but it's really the interpersonal skills and

demonstration of good interactions... that make the difference", says Padgett.

Edwards says that the environment at Monsanto today is very different from when she first joined, when essentially "you did what your boss told you to do". Today, the company is more interested in hiring people who are capable of bringing their ideas into the team, spotting an opportunity and capitalizing on it, as well as those who respond well to criticism or outside suggestion in ways that improve their research. Just being an independent person who works very hard and has the thrill of discovery does not necessarily fit with the needs of a company like Monsanto.

As well as belonging to a project team, each researcher also belongs to a 'functional' group of others with similar skills. Project teams have the goal of getting their product to market and comprise members of the different functional groups. Thus people's business and strategic thinking is developed in parallel with their functional skills, says Edwards.

Jay Pershing, Monsanto's team leader in entomology, finds that individuals are more accountable for their actions, and are more committed when working in teams. He says that for the team approach to be successful, management must accept that teams can make decisions and must offer recognition and rewards where appropriate. In the more hierarchical environment of the past, much of the credit for accomplishments made by the researchers was given to the lab chief, says Pershing.

Another recent move by Monsanto is to reduce the hierarchy of the corporate structure. About five years ago, everyone was asked to write their own job title, which should simply describe what they do. As a consequence, rank and order was minimized. The company's organizational chart today is said to resemble a honeycomb more than a typical corporate chart, where individual cells interface at areas of common research or business interest.

Joyce Fry, a project leader in monocot transformations, and Pershing, are good examples of how the reformed system allows people to progress within Monsanto irrespective of academic qualification. Fry joined Monsanto in 1967 straight after her degree in medical technology to carry out exploratory biochemistry. She moved with her supervisor to the cell biology group in the early 1970s and began working on plant protoplasts at the cellular level. She was also trying to select plants resistant to Roundup herbicide. In 1979, when the crop-transformation group was formed, she moved into the corporate research laboratories. "That's when all of the growth started", she says, "particularly in the area of genetic engineering". With her work in the crop-transformation group, she has switched from crop to crop to further this technology. Although Fry does not have a PhD, she is now project leader in monocot transformations, as well as the author of many patents.

Similarly, Pershing has risen to a position of leadership within the organization. With a masters degree in entomology, Pershing says that "I have never let my degree level affect what I do at Monsanto". There is no doubt that the flattening of the organizational structure and the de-emphasis on titles have helped to eliminate artificial barriers, Pershing says. He adds that this is key if both the company and the individual are to reach full potential. Edwards concurs and says that at Monsanto "the degree isn't the hallmark and it's not the measuring stick". Eventually, she adds, you can have the equivalent of a PhD by experience and education "on the job". This can present a problem when recruiting from the academic community, where rank and title carry considerable weight.

Another aspect of the new Monsanto culture is the 'open market' system, allowing individuals to move laterally to other projects. "When I first joined here... if you wanted to move from your current boss to another

Table 1 Breakdown of 1998 funding for the US Department of Agriculture (USDA)

Agricultural Research Service (ARS; principal in-house physical and biological science research agency in USDA)	\$741 million (President's request)
Economic Research Service (ERS; principal intramural social science research agency in USDA)	\$54 million (President's request)
Cooperative State Research, Education, and Extension Service (CSREES), which is the federal partner in the USDA-supported system of extramural scientific research, higher education and extension in the US	\$424 million (President's request for the Research and Education section of CSREES)
Intramural (ARS and ERS)	\$796 million
Extramural (CSREES – Research and Education)	\$424 million
Approximate total of extramural funds that are competitively awarded	\$132 million (\$130 million for the National Research Initiative and \$2 million for a new programme of competitive research grants on food safety)

er project, and if it was not a promotion, your current boss had to okay the move", says Edwards. Now, employees have the freedom to apply for any position they like, she says.

Loosening the purse strings

The extramural funding of academic research in agricultural science, especially in the age of molecular technology, does not begin to compare with that of medical science and extramural funding by government agencies like the National Institutes of Health (NIH). Table 1 shows a breakdown of 1998 funding for the United States Department of Agriculture. (See Table 2 for Web sites.)

Bruno Sobral, group leader for plant and microbial genomics at the National Center for Genomic Resources (NCGR) in Santa Fe, New Mexico, notes that the average award for competitive grants is about \$50,000 with a limit of 2–3 years, and that top awards are generally limited to about \$100,000, again lasting only 2–3 years. It is hard to imagine research using molecular techniques getting very far at this level of funding.

Donald Nuss, director of the University of Maryland's Center for Agricultural Biotechnology (CAB), says that the application of molecular-biology techniques has significantly accelerated the pace of discovery in agricultural research. Many key findings of relevance to agriculture are increasingly appearing in highly visible, non-specialized journals, performed in non-traditional academic departments and institutes and funded by non-traditional sources, he says. Although Nuss thinks that the rate of progress is increasing, limitations in the current funding structures for agriculturally related molecular biology are hindering the field from attaining its potential.

Nuss concludes that government agencies, the USDA in particular, have a golden opportunity to strengthen existing extramural research support for agricultural molecular biology to benefit agriculture in the same way that the NIH extramural

grants programme has benefited medicine. Serious debate is required to determine how current funding structures can be revamped and strengthened, he says.

An obvious area of growth in biotechnology is in bioinformatics. In the public sector there is a temporary shortage of researchers with skills in this area, says NCGR's Sobral. Presumably, this is also true for industry. Nuss says that there is a particular need for people with bioinformatics training who also have specific knowledge of a species. One of Sobral's main goals through his work at NCGR is to make connections with genomics researchers, much as he did at Asilomar in March. Together with colleagues from the pathology departments of New Mexico State University, the University of California at Davis and Iowa State University, he proposed the development of a *Phytophthora* database, and is talking to people involved in genome initiatives, coordinating the efforts of those working on similar species and helping to eliminate redundancy. "Not everyone needs to create their own system and database from scratch," he says. Without communication and coordination, many databases may be incompatible with other systems, or organized in a way that makes them useless to anyone but the creators, he says.

In his role as a bioinformatics 'ambassador', Sobral hopes to eliminate some of these problems. He believes that the 'big bang' for genomics resources will not be the accessibility of small databases but their coordination into larger ones. Wide-ranging accessibility, perhaps using Java Internet tools, for example, allows different ideas to be explored by many different groups using the same data.

Collaboration is particularly important for small genome projects with limited financial resources. Sobral and Nuss's meeting at Asilomar was partly responsible for Nuss proceeding with a *Phytophthora infestans* expressed-sequence-tag project

at CAB despite the fact that CAB will have to develop bioinformatics resources from scratch. In terms of funding, Nuss and CAB are fortunate. Nuss's research is funded partly by NIH grants, owing to medical interest in fungal pathogens. CAB, established in 1987, also has money from the state university system. Another source is MIPs (Maryland Industrial Partners), a joint programme between the University of Maryland and industry partners, based on research that has technical merit and the potential for commercial development. Other money comes from competitive extramural grants, as well as from industry, which now provides 16–18 per cent of CAB's funding.

Monsanto's Padgett says that all companies with an interest in advanced technology in agriculture are pooling strategies for mining genomic information. For Monsanto, this strategy may involve in-house development, outsourcing, partnering or possibly purchasing companies with information technology. "We [Monsanto] are making a fairly reasonable investment in genomics technology — in hardware and software infrastructure", says Padgett. As for information specialists, he says, Monsanto does not yet have a huge group in this area. In this regard, DowElanco is taking a more conservative approach, opting instead to be the "preferred collaborator" with academic groups or companies developing information technology. Alan Gould, DowElanco's director of biotechnology and plant genetics, says that "at the moment, I see that it has a great potential to add value to the industry in terms of focusing on particular things. But will I go out and hire a genomics expert? Not for quite a few years yet."

Molecular geneticists familiar with bacterial genetics are also being hired by Monsanto, as many genes and pathways that can be manipulated can either serve as models or have mutants available in bacteria. Padgett says that Monsanto's project on biodegradable plastics in plants has reached the point where researchers can move bacterial genes to plants, getting experience with bacterial genetics, complementing mutations with a particular gene, and searching not only the DNA databases but also searching different bacterial libraries for genes.

According to Gould, microbiology and secondary products will head the next wave of developments in agricultural biotechnology. But he says there is usually a lag of at least five years from the time the market develops a need for skills to the time when universities begin to produce the people with those skills. In the early 1980s, for example, plant molecular biologists were scarce but now there is a huge over-supply, whereas, says Gould, "you have to search long and hard to find talent" in microbiology. □

Brendan Horton is in Nature's Washington office.

Table 2 Sites of Internet interest for agricultural biotechnology

Site name	URLs: http://
General	
The National Center for Genomic Resources	www.ncgr.org/gpi/odyssey/green/
USDA (Budget)	www.usda.gov/agency/obpa/Home-Page/obpa.html
Cooperative State Research, Education, and Extension Service (CSREES)	Home page at http://www.ree.usda.gov
Information on the National Research Initiative Competitive Grants Program, the USDA's main external competitive grants programme, can be accessed from the CSREES Home Page (see above; find Funding Opportunities, then all Funding Opportunities, then National Research Initiative)	
Corporations	
Monsanto	www.monsanto.com/MonPub/index.html
DowElanco	www.dowelanco.com/
Education	
Colorado State University, Dept of Entomology	www.colostate.edu/Depts/Entomology/ent.htm
University of Arizona, Dept of Entomology	ag.arizona.edu/ENTO/entohome.html
Iowa State University, Dept of Agriculture	www.ag.iastate.edu/
University of Maryland, Center for Agricultural Biotechnology	www.umbi.umd.edu/~cab/index.html

Monsanto reinvents itself after 95 years

Diane Gershon

These are changing times at Monsanto. The giant agrochemical concern, which was founded in 1901 with a \$5,000 investment by John Queeny to make saccharin, is spinning off its chemical business to form two separate, publicly traded companies. The process should be complete by the end of this year.

The decision to separate is in part a recognition that these industry sectors serve very different markets and have different research and investment requirements. The split will enable the companies to focus on new strategic initiatives and business opportunities, as well as to improve operating efficiencies. In the case of the chemicals business, it will enable the company to address the problem of increased competition, particularly from the Pacific Rim countries. The move also reaffirms the decision made 15–20 years ago by Monsanto's management to begin investing heavily in agricultural biotechnology, particularly at a time, says Robert Fraley, president of the agricultural sector, when "there was a lot of overpromise, hype and expectation for early returns that never materialized and Wall Street was pretty much turned off".

These are "two strong businesses that are going in separate directions", Fraley says. In 1996, the company's net sales were approximately \$9.3 billion, up 3 per cent on the previous year, almost two-thirds of which was on the life-science side (including agricultural products, pharmaceuticals and food ingredients). Fraley did a postdoc in biochemistry at the University of California, San Francisco, before moving to Monsanto 15 years ago. The main objective, he says, is to change the view of Monsanto from its 100-year-old mid-western roots and to transform it into what many (Fraley included) are calling the world's largest high-tech life science start-up.

In the past, Monsanto's bulk industrial chemicals business served the company well. However, the rising cost of petroleum (the basis for many of its products) in the 1970s, together with mounting pressure from environmentalists, forced the company to take a hard look at its future.

Until 1979, Monsanto's chief scientists had always been chemists by training. But that changed with the appointment that year of Howard Schneiderman, a geneticist from the University of California, Irvine. Philip Needleman, his successor and the current chief scientist, also has a life-science background. It was Dick Mahoney (Monsanto's former chairman and chief executive officer) and Schneiderman who, Fraley says, helped steer the company through its initial



Figure 1 Monsanto sees opportunities for growth in agricultural biotechnology.

investments in agricultural biotechnology in the early 1980s, and who recruited a core team of researchers — many of whom now lead the company in this area. Both can also be credited, he says, for creating the will to build the current Life Sciences Research Center in St Louis in 1984 (Figure 1), "when it was not at all reasonable or rational that that was the thing to do".

According to Gary Barton, Monsanto's director of biotechnology communications, there were 10 years after the life-science facility was built when there were no products, only promise. A major part of the decision to invest in agricultural biotechnology was predicated on the continued success of sales of flagship products, like the non-selective herbicide Roundup and the artificial sweetener NutraSweet, to keep the money flowing in during the lean years. (Monsanto's life-science facility costs about \$50 million a year to run.)

Although Mahoney and Schneiderman can take much of the credit for setting Monsanto on this new course, it is Robert Shapiro, the current chairman and CEO, who was the chief architect of the decision to split the company in two. And it will be Shapiro who will head the new life-sciences company. Shapiro is also in part responsible for the change in organizational structure at Monsanto in an effort to shed the company's hierarchical, traditional image (see pages 431–432). Many employees who can remember the old days say the atmosphere is now more relaxed, meetings less stuffy and upper management quite approachable.

The Life Sciences Research Center, which supports 1,800 employees, more than 300 of whom are PhDs, comprises four buildings: two, five-storey-high interconnected research buildings house most of the 250 laboratories. The nearby plant agricultural building also contains laboratories, as well as a soil-potting facility for reconstituting any soil type, insect-

taries, and 110 climate-controlled growth chambers, which can simulate any environment, whether it be a rice paddy in the Philippines or a Canadian summer for growing canola. Twenty-six greenhouses on top of the facility provide almost two acres under glass. The fourth building, a pilot-scale fermentation facility, contains two 1,500-litre fermenters to produce proteins for human health-care and animal nutrition projects.

The reorganization of the life-science business has involved the consolidation of existing programmes into three ongoing business sectors: agriculture, pharmaceuticals, and food and consumer products. Fraley believes, however, that the most exciting growth opportunities are going to be at the intersections between these three areas, and so the real challenge will be how to create new connections between these traditionally separate industries.

With this in mind, the company has created three new start-up business areas: nutrition, sustainable development, and health and wellbeing. In the nutrition area, for example, this will involve producing foods or food ingredients with more health-related attributes, such as soya beans and canola with a lower saturated fat content, potatoes with a higher solids content or corn and soya beans with an increased essential amino-acid content. It may sound like science fiction to call plants 'biofactories' producing biodegradable plastic polymers and other biomaterials, as well as naturally 'coloured' cotton, but research is already under way at Monsanto to engineer plants to produce products by more environmentally sustainable practices.

Over the next couple of years, David Fischhoff, director, macronutrients in Monsanto's nutrition sector, expects the company to hire well over 100 people in some of these areas, particularly those with skills in the core disciplines of biochemistry, microbiology and molecular biology. This, he says, would also include an effort to build up the company's core capabilities in areas such as genomics, bioinformatics and protein engineering, which span multiple business areas and which require an investment that no one business sector can make on its own. Building internal strengths in these key areas is not only essential to the success of any life sciences company, says Fischhoff, but it also enables the company to be a good "receptor for technologies that are developed on the outside".

Not all the new jobs are expected to be at the St Louis facility. Monsanto is bolstering its research capabilities in Europe at its facility outside Brussels, and is looking to other parts of the world, particularly China and India, to create either freestanding, Monsanto-owned research centres, or to form collaborations with local academic institutions. This will enable the company to take advantage of the regional skill base, to

tap into local networks and to gain a better understanding of the market needs in that area, says Fischhoff.

Monsanto has also recently strengthened its hand in agricultural biotechnology by investing in several biotechnology start-up companies. Last year the company acquired Agracetus of Madison, Wisconsin, from W. R. Grace Co., and last month it announced its intention to acquire outright Calgene Inc. of Davis, California, having previously owned a controlling interest in the company. Although Calgene is perhaps most well-known for the disappointing launch of its genetically engineered, rot-resistant FlavrSavr tomato, Fraley says it was the company's cotton business and its expertise in the oil-modification area that were of particular interest.

Assimilating a smaller entrepreneurial enterprise into a corporate giant like Monsanto without hamstringing the company in the process is a challenge. But Fraley says that Monsanto intends to take full advantage of the fact that these two companies have separate cultures and histories, as well as relationships with universities: furthermore, no employees left Agracetus as a result of the acquisition. Monsanto can do a lot more for these companies than just provide additional cash, he says, by offering a synergistic science base, a whole new set of relationships to the food industry, as well as expertise in commercial development and on regulatory matters. The company is already moving people from Monsanto to Wisconsin to strengthen Agracetus's programmes and capabilities, and is building more laboratory space and greenhouse facilities there.

Monsanto has had agricultural biotechnology products in the marketplace only in the past three years (recombinant bovine somatotropin was launched early in 1994). Last year, in particular, was a benchmark year on the crops side, with commercial acreages of a number of genetically altered crops planted, perhaps signalling that agricultural biotechnology has finally begun to deliver its promise.

The past 12–18 months has been a period of rapid change, not only for Monsanto but for the whole field of agricultural biotechnology. The crop-protection industry and the seed business have come together as companies (including Monsanto) have either acquired or formed alliances with seed purveyors in an effort to gain access to the germplasm used to produce the leading commercial varieties.

Fraley is confident about the future but recognizes that Monsanto (and others) has to do more to improve consumer acceptance of genetically engineered products, particularly in Europe. Now that the tools of agricultural biotechnology are established, the real challenge will be "knowing what's of value", says Monsanto's Janice Edwards. □

Diane Gershon is an assistant editor of Nature Medicine.

Europeans benefit from collaboration

Claire O'Brien

Pan-European collaborations are the order of the day among plant biotechnologists, who are reaping benefits from academic/industrial relationships. Although the traditional agribusiness industry is contracting, new opportunities are emerging from novel projects in the academic community.

Three hundred European plant biotechnologists gathered in Dublin last month to celebrate a highly successful exercise in scientist-led coordination of their research. Under the European Union's third Framework Programme (FP3), which has just finished, research contracts at 130 laboratories studying "Plant molecular genetics for an environmentally compatible agriculture" were coordinated by AMICA Science EEIG, a consortium of leading plant scientists.

"AMICA has enhanced scientific exchange between research teams throughout Europe," says Georges Freyssinet, scientific adviser at Rhône-Poulenc, Paris, France. By January of this year all but 12% of 270 scientists hired in 1994 on temporary contracts had found new research contracts or permanent positions (see Figure 2). "AMICA should continue to play a pivotal role" in EU research programmes, says Freyssinet. Ten current projects in the EC BIOTECH programme, from genome sequencing to molecular cell biology, come under AMICA's purview.

AMICA also has an agreement with a group of 24 plant biotechnology companies called the Plant Industrial Platform (PIP), founded in 1992 to facilitate technology transfer from EC-funded researchers. Scientists are educated in intellectual-property issues, and can share confidential pre-publication results with PIP members to decide whether to apply for patents.

One example of an AMICA-coordinated project recently taken up by a PIP company is drought-resistant sugar beet. Fritz Schöffl of the University of Tübingen, Germany, has identified several plant genes that can each act as a switch in the 'heat shock' response, whereby proteins are newly expressed in response to environmental stresses, including heat, cold and drought. Schöffl has now engineered an *Arabidopsis thaliana* plant that is permanently 'switched on' so that it constitutively expresses 30–40 heat shock proteins, with the intention that the plant will grow successfully in a greater variety of environments than the wild type. Planta GmbH, the biggest of Germany's plant breeding companies, will develop the technology for sugar beet.

Although only four companies participated in plant biotechnology contracts under FP3, the contracts under its successor, FP4, involve 38 commercial partners, says Arnd Hoeveler of the EC's DGXII. He adds: "Industry has understood that basic research is needed to develop the whole field... [and] biotechnology is offering them a lot of techniques they need in order to compete with Japan and the USA".

FP5, due to start in 1999, is being organized differently: 'key actions' will define the goals of research without specifying how they should be achieved or even which disciplines will be involved; this will be for scientists themselves to propose. Assuming the European Parliament and Council approve this new structure, plant scientists will need to use strong goal-oriented arguments to secure funding, says Jeff Schell of the Max Planck Institute for Plant Breeding Research (MPIZ) in Cologne, Germany, and one of AMICA's directors. However, this need not mean that blue-skies research gets ignored, because "in plant science, the distinction between fundamental and applied research is not so clear". Schell suggests plant scientists can exploit this fact when writing grant applications: "there is going to be a major difference in how [research] is presented".

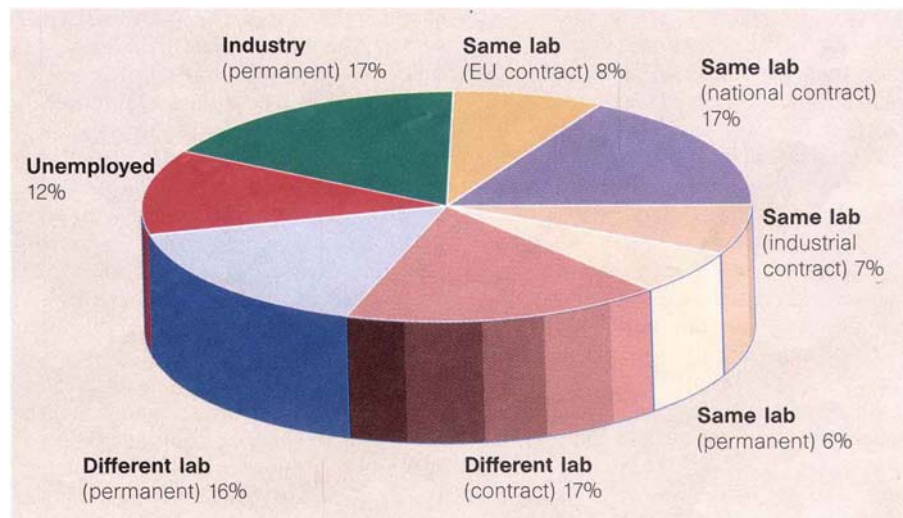


Figure 2 Current status of staff initially employed on AMICA-coordinated grants.

The mergers occurring in chemical, pharmaceutical and agriculture multinationals have dramatically reduced the number of players in the seeds and crop-protection market. In 1970, about 26 companies covered 90% of the market in crop protection (herbicides, fungicides and insecticides), but by 1995 the number was 11 and by 2005 is predicted to come down to eight. Market demands for more efficient and "environmentally friendly" products have caused a shift in emphasis away from applying chemicals to crops and towards in-built protection by plant breeding or genetic engineering.

At 15%, Novartis has the largest market share in crop-protection products, followed by Monsanto with 9.6% and Zeneca with 8.7%. All three companies are also active in research on seeds and genetically modified organisms. The 130,000 combined workforce of Ciba and Sandoz was reduced to 90,000 as a result of the merger to create Novartis, and 12% of the remaining jobs will go within two years. Of the total of 20,000 employees in the agribusiness division, 2,000 are scientists working in crop protection in Switzerland. Whereas between 30 and 50 people were recruited here each year in the past, says personnel manager Rene Frei, now only 15 are to be taken on annually. Science graduates with business or economics qualifications and experience of working abroad are favoured, as most employees learn about the business by visiting the United States or China.

AgrEvo was formed from the recent merger of the crop-protection and environmental health divisions of Hoechst and Schering, with 8,600 employees worldwide. There are about 630 employees working in research and development in Frankfurt, Germany; 400 in Chesterford Park, near Cambridge, United Kingdom; 60 in Marseilles, France and 90 in Ghent, Belgium, where AgrEvo has acquired the biotechnology company Plant Genetic Systems.

Hermann Stuebler, head of biological research, explains that the company philosophy is "there is no alternative to high-yield agriculture" if the global demand for food is to be satisfied. In achieving new methods of food production, says Stuebler, "biotechnology and chemical crop protection will, and have to, go together". Accordingly AgrEvo has developed the 'Liberty-link' concept, whereby its chemical herbicide Liberty is marketed for use with corn and canola seeds that are genetically engineered to be tolerant to it — analogous to Monsanto's Roundup herbicide and 'Roundup ready' soya bean and canola.

Before they can be marketed, products have to be tested in the field, regardless of their chemical or genetic nature. Bernhard Schreiber, head of international field development at AgrEvo, says that his department oversees experimental stations and units that work directly on farmland in over 35 countries worldwide, testing the products

Table 3 Where to find more information via the Internet

Site name	URLs: http://
Industry	
AgrEvo	www.agrevo.com/
Bayer	www.bayer.com/ usagri.bayer.com/ uscrop.bayer.com/
Novartis	www.novartis.com
Rhône-Poulenc	www.rhone-poulenc.com/
Academic institutions	
Max Planck Institute for Plant Breeding (Züchtungsforschung)	www.mpi-z-koeln.mpg.de/
Environmental pressure groups	
Greenpeace International	www.greenpeace.org
World Wide Fund for Nature	www.panda.org
Friends of the Earth	www.foe.co.uk
Others	
Consultative Group on International Agriculture Research (CGIAR)	www.cgiar.org
UK Department of Trade and Industry Biotechnology Means Business' initiative	www.dti.gov.uk/support/bmb.htm
AMICA Science EEG	www.mpi-z-koeln.mpg.de/~amica/amica00.htm
European Union BIOTECH2 programme outline	europa.eu.int/en/comm/dg12/biotech/biot3.html
European Union FAIR programme outline	europa.eu.int/en/comm/dg12/agro/agro3.html
International Pharmajobs	www.pharmajobs.com
Plant Industry Platform	www.noord.bart.nl/~biotech/pipfront.html
USDA Biotechnology Information Center	www.nal.usda.gov/bic

wherever they are sold. AgrEvo has recently built up new field-testing capacity in Poland, Czech Republic, Russia and the Ukraine. Recruiting for field units is done locally, and in growing markets such as southeast Asia, North America and Eastern Europe, employment opportunities are expanding.

"I never advertise" for new recruits, says Bayer AG's Paul Reinecke, head of the department responsible for fungicide screening and 'blue-skies' research at Monheim, Germany. Instead he telephones a network of contacts in academic laboratories in Germany and at Bayer subsidiaries in the United Kingdom, France and Italy for recommendations. Bayer's crop-protection group includes more than 1,000 research and development staff in Europe. The Monheim Agriculture Research Centre employs 350 graduates.

Rhône-Poulenc Agriculture at Ongar, United Kingdom, is developing new herbicides. The 48 plant scientists and technical staff are organized into multidisciplinary project teams, and 'floating' consultant scientists manage the application of science within their disciplines. "We particularly value 'investigative plant biochemists'", says consultant scientist David Cole, meaning plant scientists with a knowledge of chemistry and analytical techniques, as well as molecular and cell biology. However, Cole sees skills gaps: "We're not getting enough graduates coming through who can also do plant cell biochemistry and plant physiology... there's a surfeit of people offering themselves as 'spanner and screwdriver' gene cloners, but they are not so desirable". Future employees will need a broad combination of skills, a rigorous and adaptable approach,

and should be happy with automation and handling large amounts of data.

Contract research laboratories can be a good training ground for a job in industry and provide a taste of a wide variety of projects. The DLO Centre for Plant Breeding and Reproduction Research in Wageningen, The Netherlands, performs contract research for the EU and for food, chemical and plant-breeding companies, as well as foreign governments and international agencies such as the Food and Agricultural Organisation. (The DLO is the Agriculture Research Department of the Dutch Ministry of Agriculture.) The Wageningen centre is also the repository of the country's plant gene bank, and it performs the official evaluation of new cultivars to determine plant-breeders' rights; a major research activity is, of course, plant genetics. Three hundred scientists cover molecular to population biology — many with permanent posts, although the trend will be to offer only short-term contracts as the DLO will be privatized next year.

Hans Dons, deputy director, has the familiar need for "people who can deal with bioinformatics... who can handle biological data and link it with information technology and present it in the right way." The centre is also part of the consortium of European scientists sequencing *Arabidopsis*. "There's a great need for molecular biologists with knowledge of plant science", says Dons.

The DLO Institute of Agricultural and Environmental Engineering, also in Wageningen, is developing new technology for "socially acceptable" methods of production in agriculture. About 120 researchers,

with backgrounds in plant sciences or engineering, are solving such problems as how to minimize the use of herbicides to kill the leaves on potato plants, a procedure that will soon be prohibited by law throughout Europe. Other projects include ways to remove sugar beet from the ground mechanically without the accompanying soil; closed greenhouse systems for automated nutrient and water application; and robotics systems for picking cucumbers and tomatoes.

Academic research can plug some of the gaps left by the drive for industrialization. Jean Dénarié, director of research at the Laboratory of Molecular Biology of Plant-Microbe Interactions in Toulouse, France, is eager to point out that "the use of plant science to develop sustainable and ecologically sound agriculture is not always linked to the development of an industrial agriculture". For example, says Dénarié, white clover is a useful nitrogen-fixing plant to sow in pastures, but the market for these seeds is small and companies are not interested in developing or selling the product. Dénarié's institute receives funding from CNRS, the national basic research organization, and from INRA, France's applied agricultural research organization, which supports a diversity of research projects in agronomy. This is in contrast to the United Kingdom, says Dénarié, where he believes "the complexity of agriculture" receives insufficient attention due to overemphasis on genetic engineering and industry-oriented biotechnology.

"In plants, the study of signal transduction is in its infancy... this field should develop very strongly in the coming years", says Dénarié, particularly where such studies are coupled with genetics. This research could also be of huge benefit to medicine, because "plants have an incredible ability to synthesize a huge collection of secondary metabolites, such as potent antioxidants, flavonoids and alkaloids" whose synthetic pathways we do not fully understand but which are useful as therapeutics.

As the result of an initiative by the German federal government to sponsor technology-transfer programs in Cologne, Munich and Heidelberg, Heinz Saedler, acting director of MPIZ, is in the fortunate position of planning new plant research activities. He predicts the importance of "a more delicate analysis of developmental processes and the gene networks controlling them", for which researchers will need to generate "four-dimensional databases" of gene expression in space and time.

Since the late 1980s, Spain has been setting up national biotechnology centres, which include plant research. Angel Mingo-Castel of the Public University of Pamplona is setting up a centre, and describes the familiar problem of trying to attract well-trained PhDs when he is unable to offer permanent contracts. Graduates with degrees in subjects

such as plant biochemistry could alternatively look for positions in industry, but although there are new biotech companies in Spain, very few of these are involved in plant science. However, Mingo-Castel predicts that they will appear in the next five years, because researchers are producing work that needs to be taken to the market: "the technology has to go somewhere". □

Claire O'Brien is a freelance science writer.

Off-track careers

Ehsan Masood

You've got your PhD, and maybe a year or two of postdoctoral experience. Its decision time. Do you aim for traditional posts at academic institutions, look for the relatively lucrative pastures of the private sector, or even beat a path into the world of entrepreneurship? Many of the traditional employment categories are disappearing fast, opening up new, untried avenues for scientists.

The choice between industry and academic institutions, for example, is no longer a straight question of fertilizer company versus university laboratory, 'boffin' versus 'business'. The just-defeated Conservative government in the United Kingdom was particularly keen to encourage scientists to venture into the world of business through schemes such as 'Biotechnology Means Business' and 'Crusade for Biotechnology', organized by the Department of Trade and Industry in response to a long-held concern that Britain needs to learn how to capture the economic benefits of its research. The government was aware that scientists have ideas but lack commercial training (see *Nature* 381, 635; 1996). Last year it launched a £7 million (\$11 million) scheme offering advice and professional guidance on setting up small businesses.

If business is not your cup of herbal tea, yet you want something different from working in a European laboratory and don't mind taking a pay cut, you might want to head east. Most developing countries are actively engaged in plant-science research, though with a bias towards agriculture. India, Israel, Egypt and Pakistan all have strong plant-sciences sectors.

An alternative is to approach the Consultative Group on International Agriculture Research (CGIAR), an international network of research centres organized by the World Bank in Washington, DC. The CG centres were set up 25 years ago with the bold aim of helping to remove world hunger. They were at the centre of the 'green revolution' of the 1970s, when research and production of high-yielding varieties of crops, together with the liberal use of chemical fertilizers and pesticides, went some way to staving off food shortages.

Individual centres were located according to specific regional needs: for example, the International Rice Research Institute was set

up in the Philippines, the International Center for the Improvement of Wheat and Maize in Mexico, the International Crops Research Institute for the Semi Arid Tropics in Andhra Pradesh, India, and the International Center for Agriculture Research in the Dry Areas in Aleppo, Syria. Heinrich von Loesch, a spokesman for the CGIAR in Washington, says that employment information is posted on individual centres' Web sites, which can be accessed through the CGIAR main site (see Table 3).

Recently, however, the centres have become the subject of controversy. Increasing donor fatigue led last year to a long-awaited restructuring, which resulted in considerable redundancies, particularly to local auxiliary staff. Under the restructuring, centre funding is no longer guaranteed, depending more on the quality of projects that a centre can offer. Though this has drawbacks, it does favour those who can come up with relevant, practical yet innovative research projects.

The centres continue to attract opposition from environmentalist groups, who argue that high-yielding crops contribute to soil degradation. Environmentalists believe that the use of vast quantities of fertilizer is an outmoded method of cultivation. One long-standing critic of the CG system is the organization Genetic Resources Action International (GRAIN), based in Barcelona.

Organizations such as GRAIN, Greenpeace, and the World Wide Fund for Nature are useful examples of a third off-track destination for plant scientists: environmental groups. With the increase in the public profile of issues such as biotechnology in agriculture and native intellectual property rights, environmentalist pressure groups are keener than ever to employ 'experts' for scientific advice.

Duncan McLaren, senior research coordinator at Friends of the Earth, routinely employs PhD-level scientists, either on the campaigns staff or to prepare briefing documents for lobbying local and international environmental meetings. However, he points out that applicants with some background in environmental campaigns are preferred, as most of the work involves strategic planning and literature reviews rather than research. In addition, a thorough understanding of the research process is more of an advantage than a commitment to a narrow subject. "We rarely carry out primary research. But we need people who can understand the relevant literature."

Helen Wallace, head of science at Greenpeace, says that science in her organization is not an open-ended affair but is used to back up campaign messages. "In our campaigns against the release of genetically modified organisms into the environment, we benefit from the input of scientists who know that genetic manipulation is not precise, and are concerned that the risks of genetic pollution cannot be predicted". □

Ehsan Masood is a science writer for *Nature*.